

2019

## Three Essays on Nonparametric Hypothesis Testing and Stochastic Frontier Analysis

Taining Wang

West Virginia University, [tnwang@mix.wvu.edu](mailto:tnwang@mix.wvu.edu)

Follow this and additional works at: <https://researchrepository.wvu.edu/etd>



Part of the [Econometrics Commons](#)

---

### Recommended Citation

Wang, Taining, "Three Essays on Nonparametric Hypothesis Testing and Stochastic Frontier Analysis" (2019). *Graduate Theses, Dissertations, and Problem Reports*. 3930.

<https://researchrepository.wvu.edu/etd/3930>

This Dissertation is protected by copyright and/or related rights. It has been brought to you by the The Research Repository @ WVU with permission from the rights-holder(s). You are free to use this Dissertation in any way that is permitted by the copyright and related rights legislation that applies to your use. For other uses you must obtain permission from the rights-holder(s) directly, unless additional rights are indicated by a Creative Commons license in the record and/ or on the work itself. This Dissertation has been accepted for inclusion in WVU Graduate Theses, Dissertations, and Problem Reports collection by an authorized administrator of The Research Repository @ WVU. For more information, please contact [researchrepository@mail.wvu.edu](mailto:researchrepository@mail.wvu.edu).

# **Three Essays on Nonparametric Hypothesis Testing and Stochastic Frontier Analysis**

**Taining Wang**

Dissertation submitted  
to the John Chambers College of Business and Economics  
at West Virginia University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy  
in  
Economics

Feng Yao, Ph.D., Chair  
Arabinda Basistha, Ph.D.  
Jane Ruseski, Ph.D.  
Xiaoli Etienne, Ph.D.

Department of Economics

Morgantown, West Virginia  
2019

Keywords: Nonparametric Hypothesis Testing; Semiparametric Models; Stochastic  
Frontier Analysis; Debt Growth; Technical Efficiency  
Copyright 2019, Taining Wang

# ABSTRACT

## Three Essays on Nonparametric Hypothesis Testing and Stochastic Frontier Analysis

Taining Wang

The first chapter proposes a nonparametric test of significant variables in the partial derivative of a regression mean function. The test is constructed through a variation based measure of the derivative in the directions of the significant variables, with the derivative estimation through a local polynomial estimator. The test is shown to have the asymptotic null distribution and demonstrated to be consistent. The chapter further proposes a wild-bootstrap test, which exhibits the same null distribution regardless of whether the null is valid or not. Through Monte Carlo studies, the test shows encouraging finite sample performances. Through an empirical application, the test is applied to infer certain aspects of regression structures on labor's earning function.

The second chapter investigates the role of debt in the firm's production frontier and technical efficiency by employing a firm-level dataset over 1998-2007 and 1998-2013. The impact of debt on frontier is decomposed into a stand-alone neutral effect and indirect non-neutral effects, which alter the output elasticity of production inputs. The effects are estimated through a semiparametric smooth coefficient stochastic frontier model. A nonzero probability for the firms to be fully efficient is allowed, modeled as a function of debt and technical progress. The study shows that an increase in debt significantly shifts firms' frontier downward across different ownerships, regions, and industries. Foreign and private firms are more efficient, with their full efficiency probability increased by debt and technical progress. By contrast, state-owned enterprises (SOEs) and collective firms are much less efficient and their probability of being fully efficient does not increase with more debt. Furthermore, lower efficiency levels are concentrated in the central and western regions and in the mining and public utility industries.

The third chapter proposes a semiparametric additive stochastic frontier model for panel data, where inputs and environment variables can enter the frontier individually and interactively through unknown smooth functions. The inefficiency has its mean function known up to certain parameters, and influenced by its determinants that may or may not appear on the frontier. The model disentangles time invariant unobserved heterogeneities from inefficiency, which can be helpful to avoid overestimating the inefficiency level. Different from conventional stochastic frontier models, the proposed model can be identified without the distribution assumption on the composite error, and consistently estimated without suffering from the *curse of dimensionality*. Thus, a large number of interested variables for frontier or inefficiency determinants can be included, a potentially attractive feature for empirical studies. The study demonstrates the appealing finite-sample performance of the proposed estimator and two related hypotheses tests through the Monte Carlo study, and performs a world production frontier analysis with 116 countries during 2001-2013.

## Acknowledgements

Devoting my time to deeply understanding a subject of my interest has long been a dream since I was a kid. I consider myself very fortunate to have such a chance to pursue economic research toward the PhD at West Virginia University. Indeed, this dream becomes true as a result of tremendous help and supports from the most important people in my life. First of all, it is my genuine pleasure to express my deep gratitude to my advisor, Dr. Feng Yao, who has consistently provided generous help to my research. Without the clear guidance and inspiration from Dr. Yao, it is impossible for me to conduct my initial research in Nonparametric Econometrics. Dr. Yao always possesses enthusiastic attitude towards research works, which greatly encourages me to face and solve challenging works throughout the research process. With detailed explanation by Dr. Yao on technical issues, I am able to go deeper in the field of Econometrics, and see many interesting and promising topics for my future works.

I would like to extend my gratitude to my committee members, Dr. Arabinda Basistha, Dr. Jane Resuske, and Dr. Xiaoli Etienne, for their constructive feedbacks and advices on my present studies. I also gratefully thank Dr. Guillermo Franco for his generous help on High Performance Computing, which significantly facilitates many of my projects.

Last of all, I greatly thank my mother Yihong Zhang for her endless love throughout my life, and my loving fiancé Jinjing Tian for her invaluable support to my PhD life and inspiration for my research. I also thank my best friend, Tingyu Li, who always have a way to cheer me up when I face obstacles.

# Contents

<b>1</b>	<b>A Nonparametric Test of Significant Variables in Gradients</b>	<b>1</b>
1.1	Introduction . . . . .	1
1.2	A test for significant variables in gradient . . . . .	3
1.2.1	Local polynomial gradient estimate . . . . .	3
1.2.2	A variation based test . . . . .	5
1.3	Asymptotic properties . . . . .	6
1.3.1	Assumptions . . . . .	6
1.3.2	Asymptotic distribution . . . . .	7
1.4	A bootstrap test . . . . .	8
1.5	Monte Carlo study . . . . .	10
1.5.1	Bivariate Case . . . . .	11
1.5.2	Trivariate Case . . . . .	12
1.5.3	A Comparison Study . . . . .	14
1.6	Empirical Application . . . . .	20
1.7	Conclusion . . . . .	24
<b>2</b>	<b>Does High Leverage Ratio influence Chinese Firm Performance? A Semiparametric Stochastic Frontier Approach with Panel Data</b>	<b>1</b>
2.1	Introduction . . . . .	1
2.2	Empirical Methodology . . . . .	5
2.3	Data . . . . .	9
2.4	Empirical Results . . . . .	13
2.4.1	Whole Sample Analysis: 1998-2007 . . . . .	13
2.4.2	Whole Sample Analysis: 1998-2013 . . . . .	20
2.4.3	A Robustness Check on Whole Sample Analysis: 1998-2007 . . . . .	23
2.4.4	Analysis across Ownerships . . . . .	24
2.4.5	Analysis across Regions . . . . .	29
2.4.6	Analysis across Industries . . . . .	33
2.5	Conclusion . . . . .	36
<b>3</b>	<b>A Fixed Effect Additive Stochastic Frontier Model with Interactions and Distribution Free Inefficiency</b>	<b>3</b>
3.1	Estimation Procedure and Hypothesis Testing . . . . .	6
3.1.1	Semiparametric Estimation . . . . .	6
3.1.2	Hypothesis Testing . . . . .	9
3.2	Monte Carlo Simulation . . . . .	11
3.3	Empirical Application . . . . .	18
3.4	Conclusion . . . . .	28

# List of Figures

1.1	Power Curve Comparison of Additivity Tests at 5% Rejection Region . . . . .	19
1.2	Plot of $m_w(w)$ estimates against $w$ . . . . .	22
1.3	Plots of $\hat{m}_Y(Y)$ and $\hat{m}_1(Y)$ in the Partially Linear Varying Coefficient Model . . . . .	23
1.4	Plots of $\hat{m}_Y(Y)$ and $\hat{m}_1(Y)$ in the Partially Linear Varying Coefficient Model . . . . .	24
2.1	SC-ZISF Estimation from the <b>Whole</b> Sample: 1998-2007 . . . . .	14
2.2	Density of Estimated Composite Error and Posterior TE: 1998-2007 . . . . .	20
2.3	SC-ZISF Estimation from the <b>Whole</b> Sample: 1998-2013 . . . . .	21
2.4	Density of Estimated Composite Error and Posterior TE: 1998-2013 . . . . .	22
2.5	SC-ZISF Estimation with Lag of Debt Ratio from the <b>Whole</b> Sample: 1998-2007 . . . . .	23
2.6	Density of Posterior Probability and Posterior TE across ownerships . . . . .	26
2.7	Yearly Averaged Posterior Probability and PTE . . . . .	27
2.8	Density of Posterior Probability and PTE across Regions . . . . .	31
2.9	Density of Posterior Probability and PTE across Industries . . . . .	35
3.1	Simulation Function . . . . .	12
3.2	SF-AMIFE: Neutral Functions Comparison . . . . .	21
3.3	SF-AMIFE: Interaction Functions Comparison . . . . .	24
3.4	SF-AMIFE: Interaction Functions Comparison . . . . .	28

# List of Tables

1.1	Empirical Size and Power from Bivariate Regression with $d = 2$ . . . . .	13
1.2	Empirical Size and Power from Trivariate Regression with $d = 3$ . . . . .	15
1.3	Additivity Test Comparison From A Bivariate Regression . . . . .	18
1.4	Descriptive Statistics of the Dataset . . . . .	21
1.5	Empirical Results From the Dataset . . . . .	21
2.1	Data Descriptive Summary Statistics . . . . .	12
2.2	The Share of Firms with Different Ownership across Regions and Industries: 1998-2007 .	13
2.3	SC-ZISF Estimation Results from the <b>Whole</b> Sample . . . . .	16
2.4	Parametric Models Estimation Results from the <b>Whole</b> Sample . . . . .	17
2.5	Parameter Estimation Results of Lag of Debt Ratio . . . . .	24
2.6	Estimation Results of SC-ZISF Model based on <b>Ownership</b> . . . . .	25
2.7	Estimation Results of the SC-ZISF Model based on <b>Region</b> . . . . .	30
2.8	Average of PTE and Posterior Probability across 31 Provinces . . . . .	32
2.9	Estimation Results of SC-ZISF Model based on <b>Industry</b> . . . . .	34
3.1	Simulation Results of Parameters $\hat{\gamma}$ in Inefficiency Mean Function . . . . .	14
3.2	Simulation Results of Function Estimates: A RE model ( $c_0 = 0$ ) . . . . .	15
3.3	Simulation Results of Function Estimates: A FE model ( $c_0 = 1$ ) . . . . .	16
3.4	Nonparametric Tests Results ( $T = 3$ ) . . . . .	17
3.5	Data Descriptive Summary Statistics . . . . .	20
3.6	OLS FE Estimation Results . . . . .	22
3.7	Nonparametric Tests Results on Frontier Functional Form and Interactions . . . . .	26
3.8	Yearly Averaged TE and TEC Ranking 2001-2013 . . . . .	27

# Chapter 1

## A Nonparametric Test of Significant Variables in Gradients

### 1.1 Introduction

Nonparametric estimation and hypothesis test gain popularity among practitioners as they are robust to functional misspecification under less restrictive assumptions (Li and Racine (2007)). The test of significant variables has been of interest to many in regression analysis, since it is often used to support/reject an economic theory or considered for model selection (see Hart (1997) for a review of nonparametric tests). The test can be performed by the sample analog of a moment condition (Fan et al. (1996), Zheng (1996), Li and Wang (1998), Li (1999), Lavergne and Vuong (2000), Hsiao et al. (2007), Gu et al. (2007)), by comparing the difference between sums of squared residuals (Ullah (1985), Dette (1999), Fan and Li (2002)), by an  $R^2$  based statistic (Su and Ullah (2012), Yao and Ullah (2013)), or by the partial derivative estimate based bootstrap procedure (Racine (1997)). General tests of a parametric model can also be performed through the integrated squared difference between parametric and nonparametric fit as in Hardle and Mammen (1993), or likelihood ratio based test (Azzalini and Bowman (1993), Fan and Huang (2001), Hong and Lee (2009)).

However, relatively little attention has been paid to test significant variables in the partial derivative of a regression function. Racine (1997) proposes a partial derivative estimate based bootstrap test for significant variables in the regression function. Sperlich et al. (2002) propose a test based on a cross derivative for the interaction term in an additive regression model. We are not aware of works in testing significant variables in the gradients, which can provide valuable insights. First, this test can offer independent insights into the structure of the gradient functions, regarding whether the partial derivative depends on all, some, or none of the variables in the regression. For example, in applied nonparametric estimation, a common practice is to report the partial derivative estimate at the mean value of independent



variables, which can be misleading unless the gradient estimate does not change much over the support of independent variables. In estimating the stochastic frontier to evaluate the productive efficiency, for example in Yao et al. (2017), one important question arises regarding whether the marginal product of a regular input depends on computerization, an environment variable. Second, this test further allows one to infer about the structure of the regression function.<sup>1</sup> For instance, in a bivariate regression, if the gradient of one variable  $x$  does not depend on the other variable  $z$ , one can infer an additive structure in the regression. If the gradient of  $x$  does not depend on  $x$ , one can expect that  $x$  enters the model linearly, with a potential varying coefficient that can depend on the other variable  $z$ . If the gradient of  $x$  does not depend on  $x$  or  $z$ , then not only  $x$  enters the model linearly, but also there is no interaction between  $x$  and  $z$  variables, i.e., a partially linear model. Finally, as pointed out by Sperlich et al. (2002), “even though a certain test based on the estimation of a functional form is superior in detecting a general deviation from the hypothetical one, a single peak or bump can often be better detected by tests based on the derivatives.”

In this paper, we fill the gap by providing a simple nonparametric test of significant variables in the partial derivative of a regression function. Recall that a general principle of the nonparametric tests mentioned above is to compare the difference between the null (parametric or semiparametric) and alternative (nonparametric) estimates. We note that in the case of testing significant variables in the derivative, it is not clear to us how to construct the gradient estimate satisfying the null restriction such that it does not depend on some variables. We overcome this hurdle by proposing a variation based test using a local polynomial gradient estimate, checking whether the gradients exhibit variations along the direction of variables that we test in the null.

Facilitated by the local polynomial gradient estimates, we obtain a simple asymptotic null distribution, whose bias and variance depend only on the error terms in the regression model. We further show that the test is consistent. Motivated by the simple structure of the null distribution, we propose a wild-bootstrap test statistic, constructed by only resampling the estimated residuals. We demonstrate that the bootstrap test statistic exhibits the same asymptotic null distribution, no matter whether the null is valid or not. This result facilitates greatly in performing the test, since the bootstrap procedure will be valid across all different types of null hypothesis, and thus there is no need to modify either the procedure or its asymptotic justification. We further illustrate its encouraging finite sample performances through a Monte-Carlo study.

Many interesting papers have considered tests for a semiparametric structure in the regression models. For example, Lewbel (1995) proposes a moment based test of the Slutsky symmetry structure in the

---

<sup>1</sup>We assume below that the variables in the regression are significant. Otherwise, one can think of the test of significance through whether the partial derivative is zero as a special case of our proposed test.

demand functions, which are estimated with a kernel method. Chen and Fan (1999), Ait-Sahalia et al. (2001), and Delgado and Manteiga (2001) have considered testing a nonparametric/semiparametric null model, with the null estimated with a kernel method. Li et al. (2003) propose a series based test for a semiparametric null model by weighting the unconditional moment. A recent work by Korolev (2018) proposes a consistent LM type specification test for semiparametric models with an increasing number of series based unconditional moments. Our test has a different focus, that is, to test for the significant variables in the gradient function. Like the other omnibus specification tests, our test does not offer a comprehensive model selection procedure, and remains silent on how to proceed if the null is rejected. However, our test on the significant variables in the gradient function can be used to infer the structure of the regression models. In the empirical section, we illustrate with a dataset on estimating the labor demand that repeated applications of our test can shed light on the search for the refined structure of the demand.

In what follows, we present the test in section 2, the theoretical properties in section 3, a bootstrap procedure and its property in section 4, a simulation study in section 5, an empirical illustration in section 6, and conclusion in section 7. All the proofs are relegated to the Appendix in the end.

## 1.2 A test for significant variables in gradient

Let's consider a regression model

$$y_i = m(W_i) + \epsilon_i, i = 1, 2, \dots, n,$$

based on identically and independently distributed (IID) random variables  $\{W_i, y_i\}_{i=1}^n$ ,  $W_i = (x_i, z_i')' \in \mathbb{R}^d$ ,  $z_i \in \mathbb{R}^{d-1}$ , and we define the partial derivative with respect to  $x$  as  $g_x(W) \equiv \frac{\partial m(W)}{\partial x}$ . Clearly, here we can consider  $x$  to be any variable in  $W$  and we assume that  $W$  are all significant in the conditional mean function, an assumption that can be tested with any of the conditional mean significant variable tests mentioned above.

### 1.2.1 Local polynomial gradient estimate

We estimate the gradient with the popular local polynomial estimation (see Fan and Gijbels (1996), Masry (1996) for discussions on the attractive properties of local polynomials and Su and Ullah (2008) for applications in simultaneous equation models). We denote the following notation to facilitate the theoretical representation. Let  $\mathbf{j} = (j_1, \dots, j_d)$ ,  $\mathbf{j}! = j_1! \times \dots \times j_d!$ ,  $|\mathbf{j}| = \sum_{i=1}^d j_i$ ,  $W^{\mathbf{j}} = W_1^{j_1} \times \dots \times W_d^{j_d}$ ,

$\sum_{0 \leq |\mathbf{j}| \leq p} = \sum_{i=0}^p \sum_{j_1=0}^i \cdots \sum_{j_d=0}^i$ , and  $(D^{\mathbf{j}}m)(W) = \frac{\partial^{\mathbf{j}}m(W)}{\partial W_1^{j_1} \cdots \partial W_d^{j_d}}$ . A local polynomial estimator of order  $p$  is obtained from minimizing the multivariate weighted least square criterion

$$\sum_{i=1}^n (y_i - \sum_{0 \leq |\mathbf{j}| \leq p} a_{\mathbf{j}}(W_i - W)^{\mathbf{j}})^2 K\left(\frac{W_i - W}{h}\right), \quad (1.1)$$

where  $K(\cdot)$  is a nonnegative kernel function on  $\mathbb{R}^d$ , and  $h \equiv h(n)$  is a scalar bandwidth sequence that goes to zero as  $n$  goes to infinity.

The first order conditions lead to the following set of equations for  $0 \leq |\mathbf{j}| \leq p$ ,

$$\frac{1}{nh^d} \sum_{i=1}^n K\left(\frac{W_i - w}{h}\right) \left(\frac{W_i - W}{h}\right)^{\mathbf{j}} y_i \equiv t_{n,\mathbf{j}}(W) = \sum_{0 \leq |\mathbf{k}| \leq p} \hat{a}_{\mathbf{k}} h^{|\mathbf{k}|} S_{n,\mathbf{j}+\mathbf{k}}(W), \quad (1.2)$$

where  $S_{n,\mathbf{j}}(W) = \frac{1}{nh^d} \sum_{i=1}^n K\left(\frac{W_i - w}{h}\right) \left(\frac{W_i - W}{h}\right)^{\mathbf{j}}$ . Following Masry (1996), let  $N_i = \binom{i+d-1}{d-1}$  be the number of distinct  $d$ -tuples  $\mathbf{j}$  with  $|\mathbf{j}| = i$ . Arrange these  $N_i$   $d$ -tuples as a sequence in lexicographical order (with the highest priority to the last position or that  $(0, \dots, 0, i)$  is the first in the sequence and  $(i, 0, \dots, 0)$  is the last), and let  $G_i^{-1}$  to denote this one-to-one map. Then the gradient estimate is  $\hat{g}_x(W) = \hat{a}_{G_1(d)}$ , since  $x$  appears as the first element in  $W = (x, z)'$ .

The local quadratic estimation is frequently utilized for estimating the first order partial derivative (Lu (1996), Ruppert and Wand (1994)) and can be sufficient in some empirical situations considered for testing significant variables of gradient. In this case, the equation (1.1) has a simple format

$$\sum_{i=1}^n (y_i - A_0 - (W_i - W)' A_1 - \text{vech}((W_i - W)(W_i - W)')' A_2)^2 K\left(\frac{W_i - W}{h}\right), \quad (1.3)$$

where  $\text{vech}(D)$  is the half-vectorization for symmetric matrix  $D_{d \times d}$ , i.e., the  $\frac{1}{2}d(d+1)$  column vector obtained by vectorizing only the lower triangular part of  $D$ . Define  $Y = (y_1, \dots, y_n)$ ,  $\mathbf{K} = \text{diag}\{K\left(\frac{W_i - w}{h}\right)\}_{i=1}^n$ ,

$$R = \begin{bmatrix} 1 & (W_1 - W)' & \cdots & \text{vech}'((W_1 - W)(W_1 - W)') \\ \vdots & \vdots & & \vdots \\ 1 & (W_n - W)' & \cdots & \text{vech}'((W_n - W)(W_n - W)') \end{bmatrix},$$

then the gradient estimate is given explicitly as  $\hat{g}_x(W) = e_2'(R'\mathbf{K}R)^{-1}R'\mathbf{K}Y$ , where  $e_2$  is a  $(1 + d + \frac{1}{2}d(d+1)) \times 1$  vector, with its 2nd element being one, and other elements being zeros. The use of local polynomial estimation offers its generality and convenience in the theoretical consideration, illustrated more clearly through the requirements on the order of  $p$  to reduce the bias (see our Assumption A4(2))

and its discussions below). Hence, we proceed to consider a general  $p$ -th order local polynomial gradient estimation in constructing the test statistics.

### 1.2.2 A variation based test

We are interested in testing the significance of variables  $\omega$  in  $g_x(W)$ , with  $\omega \subseteq W$  and  $\omega \in \mathbb{R}^{d_1}$ , where  $1 \leq d_1 \leq d$ . Let  $\omega^c$  denote the other  $d - d_1$  variables in  $W$  and we write  $\xi(W) = \xi(\omega; \omega^c)$  for any generic function  $\xi(\cdot)$ , highlighting the  $d_1$  variables  $\omega$  subjected to testing for significance.

We assume that all  $W$  variables are significant in  $m(\cdot)$ , and express our null hypothesis as  $H_0 : \omega$  not significant in  $g_x(W)$ , or equivalently, that  $g_x(\omega; \omega^c)$ 's value does not change with  $\omega$ . It is not clear to us how to estimate  $g_x(W)$  subject to the null restriction. Denoting the density of random variable  $W$  by  $f(W)$  and the density of  $\omega$  by  $f_\omega(\omega)$ , we propose to consider the following variation based measure

$$T = \int [g_x(\omega_i; \omega_i^c) - \int g_x(\omega_j; \omega_j^c) f_\omega(\omega_j) d\omega_j]^2 f(W_i) dW_i,$$

for  $j \neq i$ . Thus, it captures the variation of  $g_x(\cdot)$  in the directions of  $\omega$ . We note that  $T$  measures the distance between  $g_x(\omega_i; \omega_i^c)$  and  $\int g_x(\omega_j; \omega_j^c) f_\omega(\omega_j) d\omega_j$  with a  $L - 2$  norm. Equivalently,

$$T = E[g_x(\omega_i; \omega_i^c) - E(g_x(\omega_j; \omega_j^c) | \omega_i^c)]^2.$$

In the special case where  $E[\{g_x(\omega_i; \omega_i^c) - E(g_x(\omega_j; \omega_j^c) | \omega_i^c)\} | \omega_i^c] = 0$ ,  $T = E[E(\{g_x(\omega_i; \omega_i^c) - E(g_x(\omega_j; \omega_j^c) | \omega_i^c)\}^2 | \omega_i^c)] = E[V(g_x(\omega_j; \omega_j^c) | \omega_i^c)]$  by Law of Iterated Expectation. Note that the conditional variance  $V(g_x(\omega_j; \omega_j^c) | \omega_i^c) \geq 0$ , and it is equal to zero only when  $g_x(W)$  does not vary with  $\omega$  under the null hypothesis. So when the null is not true,  $T = E[V(g_x(\omega_j; \omega_j^c) | \omega_i^c)] > 0$ . In general,  $E[\{g_x(\omega_i; \omega_i^c) - E(g_x(\omega_j; \omega_j^c) | \omega_i^c)\} | \omega_i^c] \neq 0$ , so  $T$  does not have the variance interpretation. However, we still have  $T = 0$  under the null, and  $T > 0$  when the null is not true. Thus, this variation based  $T$  provides a valid target measure for us to perform the significance test of the gradient.

We construct an empirical estimate of  $T$  to perform the desired test through replacing the unknown in  $T$  by its local polynomial estimates,

$$\hat{T} = \frac{1}{n} \sum_{i=1}^n [\hat{g}_x(W_i) - \frac{1}{(n-1)} \sum_{\substack{j=1 \\ i \neq j}}^n \hat{g}_x(\omega_j; \omega_j^c)]^2. \quad (1.4)$$

$\hat{T}$  estimates its population counterpart  $T$ , replacing expectations with sample averages. To simplify asymptotic analysis, we restrict  $i \neq j$ , and rely on the “leave-one-out” local polynomial gradient estimate  $\hat{g}_x(W_i)$ , where the data utilized in estimation will not include the evaluation point  $W_i$ .  $\hat{T}$  also calls for a

local polynomial gradient estimate  $\hat{g}_x(\omega_j; \omega_i^c)$ , where the estimation of  $\hat{g}_x$  will not use the  $i$ -th and  $j$ -th observations. Finally, our test statistic is presented as  $nh^{2+\frac{d}{2}}\hat{T}$ , a scaled version of  $\hat{T}$ .

### 1.3 Asymptotic properties

We characterize the asymptotic properties of our proposed test  $nh^{2+\frac{d}{2}}\hat{T}$  under the following assumptions. Below we denote a generic constant by  $C$ , the magnitude of which is inconsequential for the asymptotic analysis, and can vary from one place to another. Let's denote a generic function  $\xi(W) \in C^j$  if  $\xi(W)$  and all of its partial derivatives of order  $\leq j$  are continuous and uniformly bounded on  $\mathbb{R}^d$ .

#### 1.3.1 Assumptions

**A1.** (1)  $\{y_i, W_i, \epsilon_i\}$  is IID, and we denote the density of  $W_i = (x_i, z_i')'$  by  $f(W)$ ; (2)  $f(W) \in C^2$  and  $0 < \inf_{W \in G} f(W)$  for a compact subset  $G$  of  $\mathbb{R}^d$ . (3) The conditional density of  $W$  given  $\epsilon$  is  $f_{W|\epsilon}(W)$ , and it is uniformly bounded on  $\mathbb{R}^d$ .

**A2.**  $K(W) : \mathbb{R}^d \rightarrow \mathbb{R}$  is a product kernel  $K(W) = \prod_{j=1}^d \mathcal{K}(w_j)$  with symmetric  $\mathcal{K}(w_j) : \mathbb{R} \rightarrow \mathbb{R}$  such that: (1)  $|\mathcal{K}(w)w^j| \leq C$  for all  $w \in \mathbb{R}$  with  $j = 0, 1, \dots, 2p+1$ ; (2)  $\int |w^j \mathcal{K}(w)| dw \leq C$  for  $j = 0, 1, \dots, 2p+1$ ; (3)  $\int \mathcal{K}(w) dw = 1$ ,  $\int w \mathcal{K}(w) dw = 0$ ; (4)  $\mathcal{K}(w)$  is continuously differentiable on  $\mathbb{R}$  with  $|w^j \frac{d}{dw} \mathcal{K}(w)| \leq C$  for all  $w \in \mathbb{R}$  and  $j = 0, 1, \dots, 2p+1$ .

**A3.** (1)  $E(\epsilon|W) = 0$ ,  $E(\epsilon^2|W) = \sigma^2(W)$ ,  $\sigma^2(W) \in C^1$  and  $E(\epsilon^4|W) \in C^1$ ; (2)  $m(W) \in C^{p+1}$ ; (3)  $E|g_x(\omega_i; \omega_j^c)|^4 < C$  for all  $i$  and  $j$ .

**A4.** (1)  $nh^{d+2} \rightarrow \infty$  as  $n \rightarrow \infty$ ; (2)  $nh^{2p+\frac{d}{2}+2} \rightarrow 0$  as  $n \rightarrow \infty$ , and  $p+1 > \frac{d}{4}$ .

**A5.** Let  $\mu_{k,j} = \int_{\mathbb{R}^d} \psi^j K(\psi) d\psi$ . Define  $S_{i,j}$  to be a  $N_i \times N_j$  matrix with its  $(l, m)$  element being  $(S_{i,j})_{l,m} = \mu_{k, G_i(l)+G_j(m)}$ . Then  $S$  is a positive definite matrix such that  $S = \begin{bmatrix} S_{0,0} & S_{0,1} & \cdots & S_{0,p} \\ S_{1,0} & S_{1,1} & \cdots & S_{1,p} \\ \vdots & \vdots & \vdots & \vdots \\ S_{p,0} & S_{p,1} & \cdots & S_{p,p} \end{bmatrix}$ .

Assumption A1 requires that observations are IID across  $i$ , a typical assumption applicable to cross-sectional data. Furthermore, we need the density of  $W$  to be smooth and bounded away from zero, enabling uniform argument on the density estimator component in the local polynomial estimation. The conditional density  $f_{W|\epsilon}(W)$  is assumed to be uniformly bounded, to obtain the uniform convergence of the local linear conditional mean estimator for the bootstrap procedure in Theorem 3. A2 gives standard moment and smoothness conditions on the kernel function to be used in the local polynomial estimation (see Masry (1996) and Su and Ullah (2008)). A3 requires that the conditional heteroskedasticity function, the conditional mean function to be smooth, and the fourth moment of the gradient function (only needed

for the alternative asymptotic distribution) to be either smooth or bounded. A4 specifies the bounds on the speed at which the bandwidth approaches zero as sample sizes increase, which allows a wide range of choices of bandwidths. Specifically, A4(2) reveals the delicate requirement between the order of local polynomial and the dimension of  $W$  that are needed for the leading bias term to approach zero. We will discuss the choice of bandwidth which easily satisfies A4 in more detail in the simulation section. Finally, A5 enables us to state the probability limit of the local polynomial estimation (For example, see A1 in Su and Ullah (2008)).

### 1.3.2 Asymptotic distribution

We present the asymptotic null distribution in Theorem 1. Denote the  $(i, j)$ th element of matrix  $A$  by  $A_{i,j}$ , note that by the definition of  $G(\cdot)$  function,  $i = G_{|\mathbf{j}|}^{-1}(\mathbf{j})$ , and we define

$$SK(\Psi) = \sum_{0 \leq |\mathbf{j}| \leq p} \sum_{i=1}^{N_{|\mathbf{j}|}} S^{-1}_{1+d, \sum_{i'=0}^{|\mathbf{j}|-1} N_{i'}+i} K(\Psi) \Psi^{G_{|\mathbf{j}|}(i)} \quad (1.5)$$

**Theorem 1.** Suppose the null  $H_0$ :  $\omega$  not significant in  $g_x(W)$  holds. Define  $B_1 = (\int \sigma^2(W) dW)(\int SK^2(\Psi) d\Psi)$ ,  $B_2 = [\int \frac{\sigma^2(W_i)}{f(W_i)} f(\omega_i; \omega_i^c) f(\omega_t; \omega_j^c) f(\omega_t; \omega_l^c) dW_t d\omega_i d\omega_j^c d\omega_l^c][\int SK(\psi_{11}; \psi_{12}) SK(\psi_{21}; \psi_{12}) d\psi_{11} d\psi_{12} d\psi_{21}]$  for  $i \neq j \neq t \neq l$ ,  $B_3 = [\int \sigma^2(W_i) f(\omega_i; \omega_l^c) dW_i d\omega_l^c][\int SK(\Psi_1) SK(\psi_{21}; \psi_{12}) d\Psi_1 d\psi_{21}]$  for  $i \neq l$ , and  $\Omega = 2(\int \sigma^4(W) dW)[\int (\int SK(\Psi_1 + \Psi) SK(\Psi) d\Psi)^2 d\Psi_1]$ , then under assumptions A1(1), (2), A2, A3(1), (2), A4 and A5, as  $n \rightarrow \infty$  we have

$$nh^{2+\frac{d}{2}}[\hat{T} - (\frac{1}{nh^{2+d}} B_1 + \frac{1}{nh^{2+d-d_1}} (B_2 - 2B_3))(1 + o_p(1))] \xrightarrow{d} N(0, \Omega).$$

The asymptotic null distribution of  $\hat{T}$  in Theorem 1 provides basis for performing hypothesis test under  $H_0$ . Below we briefly sketch the idea used to analyze  $\hat{T}$ , providing insights into its asymptotic behavior.

We can write

$$\begin{aligned} \hat{T} &= \frac{1}{n} \sum_{i=1}^n [g_x(W_i) - \frac{1}{n} \sum_{l=1}^n g_x(\omega_l; \omega_i^c)]^2 + \frac{1}{n} \sum_{i=1}^n [\hat{g}_x(W_i) - g_x(W_i) - \frac{1}{n} \sum_{l=1}^n (\hat{g}_x(\omega_l; \omega_i^c) - g_x(\omega_l; \omega_i^c))]^2 \\ &\quad + \frac{2}{n} \sum_{i=1}^n [g_x(W_i) - \frac{1}{n} \sum_{l=1}^n g_x(\omega_l; \omega_i^c)] [\hat{g}_x(W_i) - g_x(W_i) - \frac{1}{n} \sum_{l=1}^n (\hat{g}_x(\omega_l; \omega_i^c) - g_x(\omega_l; \omega_i^c))] \\ &= T_{1n} + T_{2n} + T_{3n}. \end{aligned}$$

Under the null hypothesis,  $T_{1n}$  and  $T_{3n}$  vanish, so we only focus on  $T_{2n}$ . It is easy to see that

$$\begin{aligned} T_{2n} &= \frac{1}{n} \sum_{i=1}^n [\hat{g}_x(W_i) - g_x(W_i)]^2 + \frac{1}{n} \sum_{i=1}^n [\frac{1}{n} \sum_{l=1}^n (\hat{g}_x(\omega_l; \omega_i^c) - g_x(\omega_l; \omega_i^c))]^2 \\ &\quad - \frac{2}{n} \sum_{i=1}^n [\hat{g}_x(W_i) - g_x(W_i)] [\frac{1}{n} \sum_{l=1}^n (\hat{g}_x(\omega_l; \omega_i^c) - g_x(\omega_l; \omega_i^c))] \\ &= T_{21} + T_{22} + T_{23}. \end{aligned}$$

Let's comment here on the asymptotic behavior of  $T_{21}$ . We show in the Appendix (Lemma 1) that uniformly for all  $W_i \in G$ , a compact subset of  $\mathbb{R}^d$ ,

$$\begin{aligned} \hat{g}_x(W_i) - g_x(W_i) &= \frac{1}{nh^{d+1}f(W_i)} \sum_{j=1}^n SK\left(\frac{W_j - W_i}{h}\right) \left( \sum_{|\mathbf{k}|=p+1} \frac{h^{p+1}}{\mathbf{k}!} (D^{\mathbf{k}}m)(W_i + \lambda(W_j - W_i)) \left(\frac{W_j - W_i}{h}\right)^{\mathbf{k}} + \epsilon_j \right) (1 + o_p(1)). \text{ Thus,} \\ T_{21} &= \frac{1}{n^3} \sum_{i=1}^n \sum_{j=1}^n \sum_{t=1}^n \frac{1}{f(W_i)^2 h^{2d+2}} SK\left(\frac{W_j - W_i}{h}\right) SK\left(\frac{W_t - W_i}{h}\right) [\epsilon_j \epsilon_t \\ &\quad + \epsilon_j \sum_{|\mathbf{k}|=p+1} \frac{h^{p+1}}{\mathbf{k}!} (D^{\mathbf{k}}m)(W_i + \lambda(W_t - W_i)) \left(\frac{W_t - W_i}{h}\right)^{\mathbf{k}} + \epsilon_t \sum_{|\mathbf{k}|=p+1} \frac{h^{p+1}}{\mathbf{k}!} (D^{\mathbf{k}}m)(W_i + \lambda(W_j - W_i)) \left(\frac{W_j - W_i}{h}\right)^{\mathbf{k}} \\ &\quad + \sum_{|\mathbf{k}|=p+1} \frac{h^{p+1}}{\mathbf{k}!} (D^{\mathbf{k}}m)(W_i + \lambda(W_t - W_i)) \left(\frac{W_t - W_i}{h}\right)^{\mathbf{k}} \sum_{|\mathbf{k}|=p+1} \frac{h^{p+1}}{\mathbf{k}!} (D^{\mathbf{k}}m)(W_i + \lambda(W_j - W_i)) \left(\frac{W_j - W_i}{h}\right)^{\mathbf{k}}] \\ &= T_{211} + T_{212} + T_{213} + T_{214}. \end{aligned}$$

In the proof of Theorem 1, we demonstrate that the asymptotic behavior of  $T_{21}$  above is totally determined by  $T_{211}$ . Thus, the expressions based on the bias of estimation  $(D^{\mathbf{k}}m)$  do not play a role asymptotically. Specifically, we can show that

$$\begin{aligned} nh^{2+\frac{d}{2}}(T_{211} - \frac{1}{nh^{2+d}}B_1) &= nh^{2+\frac{d}{2}} \begin{pmatrix} n \\ 2 \end{pmatrix}^{-1} \sum_{i=1}^n \sum_{j=1}^n \frac{\epsilon_i \epsilon_j}{h^{2+2d}} \int SK\left(\frac{W_j - W_t}{h}\right) SK\left(\frac{W_i - W_t}{h}\right) \frac{1}{f(W_t)} dW_t (1 + o_p(1)) \\ &\xrightarrow{d} N(0, \Omega). \end{aligned}$$

Thus, the quadratic form of the innovations  $\{\epsilon_i\}$  plays the dominating role of determining the asymptotic distribution (an observation also appears in Hardle and Mammen (1993) and Kress et al. (2008)). Similar arguments apply to terms  $T_{22}$  and  $T_{23}$ , and we can show that  $T_{22} + T_{23} = \frac{1}{nh^{2+d-d_1}}(B_2 - 2B_3)(1 + o_p(1))$ . Once again, only the expressions based on  $\epsilon$  contribute to the asymptotic bias terms. This observation greatly simplifies the asymptotic distribution expressions, and further motivates our proposal for the bootstrap procedure in the next section.

Under the alternative,  $g_x(\omega; \omega^c)$  varies with  $\omega$ , thus  $T_{1n}$  and  $T_{2n}$  do not vanish. We can show that  $T_{1n} = E[g_x(W_i) - E(g_x(\omega_i; \omega_i^c) | \omega_i^c)]^2 + o_p(1)$  and  $T_{3n} = o_p(1)$ . It leads to our Theorem 2, which establishes the global consistency of our test  $nh^{2+\frac{d}{2}}\hat{T}$ .

**Theorem 2.** *Suppose the alternative  $H_A$ :  $\omega$  significant in  $g_x(W)$  holds. Then under assumptions A1(1),(2), A2, A3(1)-(3), A4 and A5, we have  $\hat{T} = E[g_x(W_i) - E(g_x(\omega_i; \omega_i^c) | \omega_i^c)]^2 + o_p(1)$ . So we have as  $n \rightarrow \infty$ ,  $P(nh^{2+\frac{d}{2}}\hat{T} > c_n) \rightarrow 1$  for any positive constant  $c_n = o(nh^{2+\frac{d}{2}})$ . Thus, the test  $nh^{2+\frac{d}{2}}\hat{T}$  is consistent.*

## 1.4 A bootstrap test

The asymptotic distributions of our test  $nh^{2+\frac{d}{2}}\hat{T}$  obtained in the last section provide guidances to perform the test. For example, one can construct estimates of the unknowns in the asymptotic distribution,  $B_1, B_2, B_3$ , and  $\Omega$ , compare the standardized  $nh^{2+\frac{d}{2}}\hat{T}$  with the critical value from a standard normal

distribution and draw conclusions. However, many papers have revealed that the asymptotic normal approximation performs poorly in finite sample settings. Specifically, the consistent nonparametric test often suffers from substantial finite sample size distortions, as the distribution of the nonparametric test statistic approaches asymptotically the normal distribution at a slow convergence rate (e.g., Hardle and Mammen (1993), Li and Wang (1998), Fan et al. (2006), Hsiao et al. (2007), Gu et al. (2007)), or the approximation from the first order asymptotic theory is far too crude to be useful in practice unless the sample size is tremendously large (Hjellvik et al. (1998)). Therefore, we provide a wild bootstrap test as a viable alternative for approximating the finite sample null distribution of the test statistic.

Many construct the bootstrap test by imposing the null restriction in the bootstrap sample (i.e., Li and Wang (1998), Fan et al. (2006), Hsiao et al. (2007), Gu et al. (2007)), such that even when the null hypothesis is not true, one still obtains the null distribution of the test statistics. This is a strategy that we could have followed. For example, when  $\omega$  in  $g_x(\omega; \omega^c)$  is  $z$ , one can infer the additive structure in  $m(\cdot)$ . We can estimate  $m(\cdot)$  additively, then use the additive  $m(\cdot)$  estimate to construct the bootstrap sample. We note first that our null hypothesis is with regard to significant variables in  $g_x(\cdot)$ . Hence, placing inferred structural restrictions on  $m(\cdot)$ , though feasible, is not direct. Second, this strategy needs to be modified when  $\omega$  changes, thereby requiring one to change the argument significantly depending on what appears in  $\omega$ .

Guided by the analysis of the null distribution of the test  $nh^{2+\frac{d}{2}}\hat{T}$ , we focus on the terms involving only the innovations  $\{\epsilon_i\}$  mentioned in the last section. They have an unvarying form for all types of null hypothesis considered in this paper, thus there is no need to modify either the bootstrap procedure or its argument for its asymptotic validity. As commented before, they play the dominating role of determining the asymptotic null distribution. We will show that they are asymptotically equivalent to the test statistics under the null hypothesis. It means that we practically bootstrap from a population that always reflects the null hypothesis.

We follow Hardle and Mammen (1993) to adopt a wild bootstrap procedure. Let  $\hat{\epsilon}_i = y_i - \hat{m}(W_i)$ , where a variety of estimates can be used for the conditional mean  $m(\cdot)$ , and here we focus on the local linear estimate  $\hat{m}(W_i)$  due to its desirable properties demonstrated in Fan and Gijbels (1996). The bootstrap test contains the following steps:

Step 1: generate  $\epsilon_i^*$  as the wild bootstrap error. For example,  $\epsilon_i^*$  is generated independently from the two point distribution  $\hat{F}_i$  such that  $\epsilon_i^* = a\hat{\epsilon}_i$  for  $a = \frac{1-\sqrt{5}}{2}$  with probability  $p = \frac{\sqrt{5}+1}{2\sqrt{5}}$ , and  $\epsilon_i^* = b\hat{\epsilon}_i$  for  $b = \frac{1+\sqrt{5}}{2}$  with probability  $1-p$ . It is called the wild bootstrap error because we use only single residual  $\hat{\epsilon}_i$  to estimate the conditional distribution of  $\epsilon_i$  given  $W_i$  by  $\hat{F}_i$ . It does not mimic the iid structure of  $\{W_i, y_i\}_{i=1}^n$ . Let  $E^*(\cdot) = E(\cdot | \{W_i, y_i\}_{i=1}^n)$  be the expectation under the bootstrap distribution, i.e.,



the conditional distribution given  $\{W_i, y_i\}_{i=1}^n$ . It is easy to verify that  $E^*(\epsilon_i^*) = 0$ ,  $E^*(\epsilon_i^*)^2 = \hat{\epsilon}_i^2$ , and  $E^*(\epsilon_i^*)^3 = \hat{\epsilon}_i^3$ .

Step 2: construct the bootstrap test statistic

$$\hat{T}^* = \frac{1}{n} \sum_{i=1}^n [\hat{g}_x^*(W_i) - \frac{1}{(n-1)} \sum_{\substack{j=1 \\ i \neq j}}^n \hat{g}_x^*(\omega_j; \omega_i^c)]^2,$$

where  $\hat{g}_x^*(W_i)$  and  $\hat{g}_x^*(\omega_j; \omega_i^c)$  are calculated with the bootstrap sample  $\{W_i, \epsilon_i^*\}_{i=1}^n$ . That is, we use  $\epsilon_i^*$  as the bootstrap dependent variable.

Step 3: repeat above two steps B times, with B a large number of choice. Then the B bootstrap test statistic  $\hat{T}^*$  yields the empirical distribution of the bootstrap statistics, which is then used to approximate the finite sample null distribution of  $\hat{T}$ . The empirical p-value is obtained as the percentage of the number of times that  $\hat{T}^*$  exceeds  $\hat{T}$  in the B repetitions.

Theorem 3 below provides asymptotic justification for the bootstrap procedure above.

**Theorem 3.** *With assumptions A1(1)-(3), A2, A3(1)-(3), A4 and A5, we have as  $n \rightarrow \infty$ ,*

$$nh^{2+\frac{d}{2}} [\hat{T}^* - (\frac{1}{nh^{2+d}} B_1 + \frac{1}{nh^{2+d-d_1}} (B_2 - 2B_3))(1 + o_p(1))] \xrightarrow{d} (0, \Omega),$$

*conditionally on  $\{W_i, y_i\}_{i=1}^n$ , where  $B_1, B_2, B_3$  and  $\Omega$  are the same as given in Theorem 1.*

It indicates that the bootstrap provides an asymptotic valid approximation to the null limiting distribution of  $nh^{2+\frac{d}{2}} \hat{T}$ . Theorem 3 holds regardless of whether  $H_0$  is true or not. When  $H_0$  is true, the bootstrap procedure will lead asymptotically to the correct size of the test, since  $nh^{2+\frac{d}{2}} \hat{T}^*$  converges in distribution to the same limiting distribution under  $H_0$  as in Theorem 1. When  $H_0$  is false,  $nh^{2+\frac{d}{2}} \hat{T}$  will converge to infinity as shown in the proof of Theorem 2, but asymptotically the bootstrap critical value is still finite for any significance level  $\alpha$  different from 0. Thus  $P(nh^{2+\frac{d}{2}} \hat{T} > nh^{2+\frac{d}{2}} \hat{T}^*) \rightarrow 1$ , thus the bootstrap method is consistent.

## 1.5 Monte Carlo study

We consider three sets of Monte Carlo studies to demonstrate the finite-sample performance of our bootstrap test statistic  $\hat{T}^*$ . Before that, we discuss the important issue for the choice of bandwidths, since the performance of the test depends heavily on the bandwidths. We note that in the kernel based test for the semiparametric null model, it is common to utilize different bandwidths for the testing and for the estimation step, i.e., see Gozalo and Linton (2001), Sperlich et al. (2002) and Wang and Carriere

(2011). Different orders of the bandwidths are utilized to derive the asymptotic distribution of the test, such that the impact of the estimation step can be properly controlled. Sperlich et al. (2002) and Wang and Carriere (2011) also suggest a double-bandwidth strategy in estimating different components in the additive part, and it is a key tactic for achieving an increased finite-sample precision of the estimate. Our test has a different focus to test for significance variables in the gradient, and we only need one set of bandwidths to implement our test. Our assumptions A4(1) and (2) specify a relatively wide range of the bandwidths. To be specific, the order of the optimal bandwidth for regression purposes, i.e.,  $h = O(n^{-\frac{1}{2p+d}})$ , satisfies the assumptions for  $d < 4$ . Thus, the estimation based optimal criteria such as cross-validation can serve as the tuning strategy. For  $d > 4$ , a slightly undersmoothed bandwidth should be utilized.

### 1.5.1 Bivariate Case

The first study considers a bivariate regression with  $W = [x, z_1]'$ , and thus  $d = 2$ . We consider three simple null hypotheses that are satisfied by three popular structured regression models. The first null, denoted by Case 1, is that  $\omega = z_1$  is insignificant in  $g_x(W)$ . It is easy to infer that the regression model is additive, i.e.,  $m(W) = m_1(x) + m_2(z_1)$ ; The second null, Case 2, states that  $\omega = x$  is insignificant in  $g_x(W)$ , so the regression model reveals a varying coefficient structure, i.e.,  $m(W) = xm_1(z_1) + m_2(z_1)$ ; The third null, Case 3, is that  $\omega = [x, z_1]'$  is insignificant in  $g_x(W)$ , which corresponds to a partially linear model  $m(W) = x\beta + m_1(z_1)$ . We consider the following three data-generating processes (DGPs) for  $i = 1, \dots, n$ :

$$DGP_1 : y_i = 0.5 + x_i + \delta x_i^2 + z_{1i} + z_{1i}^2 + \delta_1 x_i z_{1i} + \epsilon_i$$

$$DGP_2 : y_i = 5 + 2x_i - \delta e^{1.1x_i} + z_{1i}^3 + 2\delta_1 x_i \sin(z_{1i}) + \epsilon_i$$

$$DGP_3 : y_i = 1 + x_i + \delta x_i^3 + 0.4z_{1i}^2 - \delta_1 x_i e^{z_{1i}} + \epsilon_i$$

where  $x_i$  and  $z_{1i}$  are each IID and drawn independently from a uniform distribution  $U(-2, 2)$ , and  $\epsilon_i \sim N(0, 1)$  is the error term. With a nonzero  $\delta$ , the three DGPs exhibit nonlinearity in  $x$ . With a nonzero  $\delta_1$ , we introduce an interaction term between  $x$  and  $z_1$ , in which the impact of  $x$  is linear through the three DGPs, and the impact of  $z_1$  is linear only in  $DGP_1$ .  $DGP_2$  contains a high frequency function ( $\sin(z_1)$  in the interaction), and is a modified version from Wang and Carriere (2011) which tests additivity.  $DGP_3$  is adapted from Yang et al. (2006), which test for a constant coefficient against a varying coefficient model.

We investigate the size and power of our test under Cases 1-3 with different choices of  $(\delta, \delta_1)$ . For

all three DGPs in Case 1, we investigate the size performance by letting  $\delta_1 = 0$ , i.e.,  $x$  is not interactive with  $z_1$  in  $H_0$ . We simply set  $\delta = 1$  to allow for nonlinearity in  $x$ . In Case 2, we examine the size by letting  $\delta = 0$ , i.e.,  $x$  enters the model linearly in  $H_0$ . We simply set  $\delta_1 = 1$  to allow for the presence of interaction effects. In Case 3, we set  $(\delta, \delta_1) = (0, 0)$ , i.e.,  $x$  enters the model linearly with no interaction with  $z_1$  in  $H_0$ . Different values of  $\delta_1$ ,  $\delta$  and  $(\delta, \delta_1)$  other than those chosen above allow us to explore the power performance. Here, we simply illustrate the empirical power performance by letting  $\delta_1 = 1$  in Case 1,  $\delta = 1$  in Case 2, and  $(\delta, \delta_1) = (1, 1)$  in Case 3 to save space.

We perform 500 simulations, and in each we construct the wild bootstrap test based on 299 repetitions. We utilize the Gaussian kernel function  $k(v) = \frac{1}{\sqrt{2\pi}} e^{-0.5v^2}$ , and choose a *rule-of-thumb* bandwidth  $h_\xi = C\hat{\sigma}_\xi n^{-\frac{1}{2p+d}}$ , where  $C$  is the scaling factor and  $\hat{\sigma}_\xi$  is the sample standard deviation of the variable  $\xi$ , which is either  $x$  or  $z_1$ . We set  $C = 1.0$  in our study and consider three sample sizes 50, 100, and 200.

Table 1.1 summarizes the simulation results in terms of the empirical rejection frequency from the first study, for the significant levels  $\alpha = (0.10, 0.05, 0.01)$ . For the three cases, the tests are generally oversized in smaller sample sizes under  $DGP_1$  and  $DGP_3$ , but undersized in  $DGP_2$ . As the sample size increases, the size of the test generally improves toward its nominal level across all DGPs and three cases. For the chosen parameters, the empirical power of the test in Cases 1-3 rises quickly toward unity as the sample size increases, with that in Case 1 and  $DGP_2$  increasing in a moderately slower rate. Across all Cases and DGPs, the power reaches one when  $n = 200$ , indicating that our test is consistent as claimed in Theorem 2.

### 1.5.2 Trivariate Case

The second study explores a multivariate regression with  $W = [x, Z]'$ ,  $Z = [z_1, z_2]'$ , and thus  $d = 3$ . Similar to the first study, we are interested in testing the null hypothesis that the insignificant variables in  $g_x(W)$  are  $\omega$ , where  $\omega = Z$  in Case 1,  $\omega = x$  in Case 2, and  $\omega = [x, Z]'$  in Case 3. In addition, the trivariate regression model allows us to investigate an alternative additive structure by testing the null that the insignificant variable in  $g_x(W)$  is  $\omega = z_s$  which we denote by Case 1.1, and an alternative varying coefficient structure with  $\omega = [x, z_s]'$  being insignificant in  $g_x(W)$  which we denote by Case 2.1, for  $s = 1, 2$ . Correspondingly, the null regression structure is expected to be either  $m(W) = m_1(x, Z_{-s}) + m_2(Z)$ , an overlapping additive model for Case 1.1, or  $m(W) = xm_1(Z_{-s}) + m_2(Z)$ , an overlapping varying coefficient model for Case 2.1, with  $Z_{-s}$  denoting the variables in  $Z$  excluding  $z_s$ . To accommodate the nonlinearity of  $x$  and its interaction with  $z_1$  and  $z_2$ , we consider the following:

Table 1.1: Empirical Size and Power from Bivariate Regression with  $d = 2$ 

Case 1		$H_0$ : Additive model ( $\delta = 1$ )								
	$\delta_1$	$DGP_1$			$DGP_2$			$DGP_3$		
		n=50	100	200	n=50	100	200	n=50	100	200
$\alpha = 0.10$	0.0	0.136	0.088	0.092	0.062	0.068	0.076	0.070	0.078	0.080
	1.0	0.984	1.000	1.000	0.938	1.000	1.000	0.942	1.000	1.000
$\alpha = 0.05$	0.0	0.060	0.046	0.050	0.028	0.036	0.042	0.030	0.044	0.060
	1.0	0.974	1.000	1.000	0.882	0.994	1.000	0.916	1.000	1.000
$\alpha = 0.01$	0.0	0.018	0.014	0.008	0.007	0.009	0.012	0.005	0.008	0.012
	1.0	0.908	1.000	1.000	0.798	0.988	1.000	0.838	0.996	1.000
Case 2		$H_0$ : Varying coefficient model ( $\delta_1 = 1$ )								
	$\delta$	$DGP_1$			$DGP_2$			$DGP_3$		
		n=50	100	200	n=50	100	200	n=50	100	200
$\alpha = 0.10$	0.0	0.148	0.114	0.107	0.072	0.084	0.142	0.170	0.102	0.100
	1.0	1.000	1.000	1.000	0.968	1.000	1.000	1.000	1.000	1.000
$\alpha = 0.05$	0.0	0.084	0.062	0.054	0.032	0.048	0.054	0.094	0.068	0.056
	1.0	1.000	1.000	1.000	0.958	1.000	1.000	0.996	1.000	1.000
$\alpha = 0.01$	0.0	0.038	0.024	0.016	0.006	0.008	0.014	0.018	0.024	0.013
	1.0	0.998	1.000	1.000	0.936	1.000	1.000	0.986	0.968	1.000
Case 3		$H_0$ : Partially linear model								
	$\delta = \delta_1$	$DGP_1$			$DGP_2$			$DGP_3$		
		n=50	100	200	n=50	100	200	n=50	100	200
$\alpha = 0.10$	0.0	0.168	0.140	0.100	0.122	0.108	0.096	0.154	0.116	0.104
	1.0	1.000	1.000	1.000	0.982	1.000	1.000	1.000	1.000	1.000
$\alpha = 0.05$	0.0	0.088	0.062	0.058	0.100	0.044	0.046	0.082	0.076	0.068
	1.0	1.000	1.000	1.000	0.974	1.000	1.000	1.000	1.000	1.000
$\alpha = 0.01$	0.0	0.032	0.024	0.011	0.032	0.008	0.008	0.032	0.026	0.014
	1.0	0.996	1.000	1.000	1.000	1.000	1.000	0.998	1.000	1.000

*Note:* Empirical size and power are calculated based on 500 simulations with 299 bootstrap repetitions. The *rule of thumb* bandwidths have a scaling factor  $C = 1.0$ , and  $\alpha$  is the significant level.

$$DGP_4 : y_i = 0.5 + x_i + \delta x_i^2 + \delta_1 x_i z_{1i} + \delta_2 x_i z_{2i} + z_{1i}^2 + z_{2i}^2 + z_{1i} z_{2i} + \epsilon_i$$

$$DGP_5 : y_i = 5 + 2x_i - \delta e^{1.1x_i} + 2\delta_1 x_i \sin(z_{1i}) + \delta_2 x_i \cos(-z_{2i}) + z_{1i}^3 + z_{2i}^3 + \epsilon_i$$

$$DGP_6 : y_i = 1 + x_i + \delta x_i^3 - \delta_1 x_i e^{z_{1i}} + \delta_2 x_i \cos(\pi z_{2i}) + 0.4(z_{1i}^2 + z_{2i}^2) + \epsilon_i,$$

where  $z_{2i}$  is IID and generated from  $U(-2, 2)$ , and all other variables are generated as in the first study. Note that  $\delta$  controls for the degree of nonlinearity of  $x$ ,  $\delta_1$  for the interaction between  $x$  and  $z_1$ , and  $\delta_2$  for the interaction between  $x$  and  $z_2$ . Note that  $z_s$  can be either  $z_1$  or  $z_2$  in Cases 1.1 and 2.1, and to save space, we only focus on  $z_s = z_1$  below for illustrations. Under  $DGP_{4-6}$ , we investigate the size by setting  $(\delta, \delta_1, \delta_2) = (1, 0, 0)$  in Case 1,  $(\delta, \delta_1, \delta_2) = (1, 0, 1)$  in Case 1.1 ( $\omega = z_1$ ),  $(\delta, \delta_1, \delta_2) = (0, 1, 1)$  in Case 2,  $(\delta, \delta_1, \delta_2) = (0, 0, 1)$  in Case 2.1 ( $\omega = [x, z_1]'$ ), and  $(\delta, \delta_1, \delta_2) = (0, 0, 0)$  in Case 3. We explore the power performance by simply setting  $(\delta, \delta_1, \delta_2) = (1, 1, 1)$  in each case.

We implement our bootstrap test in a similar fashion as in the first study, and summarize the simulation results in Table 1.2. Due to the *curse of dimensionality*, we expect more distorted size and power performance relative to the first study. Indeed, for the small sample with  $n = 50$  and for Cases 1 and 2, we find that the test in the trivariate  $DGP_4 - DGP_6$  in Table 2 exhibits smaller empirical power and its size is farther away from the nominal level, relative to the corresponding bivariate DGPs in Table 1. However, the test for Case 3 seems to be affected less by the *curse of dimensionality*. Focusing just on Table 2, we observe that the size throughout the five cases is overestimated with a small sample size  $n = 50$ , except in  $DGP_5$  for Case 2.1, but the size improves rapidly towards the nominal level as the sample size increases. The power approaches one quickly as the sample size increases. The large sample results are still reasonably satisfactory. For  $n = 200$ , the size of the test is fairly close the the target nominal level and the power is almost one.

### 1.5.3 A Comparison Study

Since our test statistic  $\hat{T}^*$  for the null  $\omega = Z$  can be used to infer an additive structure in the regression model, in our last study we compare the performance of  $\hat{T}^*$  with two kernel-based tests for additivity in the literature. Recall that a purely additive model is  $y = m_{add}(W) + \epsilon$ , with  $W = [W_1, \dots, W_d]'$  and  $m_{add}(W) = \mu + \sum_{j=1}^d m_j(W_j)$ . The identification conditions are  $E(m_j(W_j)) = 0$  for  $j = 1, \dots, d$ , so that  $\mu = E(y)$ .

We first consider a recent additivity test by Wang and Carriere (2011) (WC hereafter) in a cross-sectional set-up. They recognize that the conventional additivity test in Gozalo and Linton (2001),

Table 1.2: Empirical Size and Power from Trivariate Regression with  $d = 3$ 

Case 1		$H_0$ : Additive model ( $\delta = 1$ )								
		$DGP_4$			$DGP_5$			$DGP_6$		
	$\delta_1 = \delta_2$	n=50	100	200	n=50	100	200	n=50	100	200
$\alpha = 0.10$	0.0	0.160	0.104	0.092	0.160	0.134	0.117	0.198	0.170	0.146
	1.0	0.966	1.000	1.000	0.818	0.934	1.000	0.902	1.000	1.000
$\alpha = 0.05$	0.0	0.092	0.088	0.062	0.106	0.076	0.042	0.130	0.098	0.064
	1.0	0.940	1.000	1.000	0.732	0.900	0.920	0.886	0.998	1.000
$\alpha = 0.01$	0.0	0.048	0.020	0.012	0.078	0.036	0.006	0.052	0.046	0.035
	1.0	0.886	0.994	1.000	0.600	0.900	0.944	0.768	0.896	0.970
Case 1.1		$H_0$ : Overlapping additive model ( $\delta = \delta_2 = 1$ )								
		$DGP_4$			$DGP_5$			$DGP_6$		
	$\delta_1$	n=50	100	200	n=50	100	200	n=50	100	200
$\alpha = 0.10$	0.0	0.106	0.120	0.110	0.206	0.012	0.080	0.166	0.110	0.080
	1.0	0.920	1.000	1.000	0.846	0.970	1.000	0.906	1.000	1.000
$\alpha = 0.05$	0.0	0.042	0.072	0.042	0.112	0.074	0.034	0.106	0.060	0.042
	1.0	0.850	0.996	1.000	0.782	0.940	0.998	0.862	0.976	1.000
$\alpha = 0.01$	0.0	0.018	0.022	0.014	0.092	0.074	0.032	0.060	0.020	0.006
	1.0	0.694	0.986	1.000	0.684	0.940	0.942	0.778	0.898	0.998
Case 2		$H_0$ : varying coefficient model ( $\delta_1 = \delta_2 = 1$ )								
		$DGP_4$			$DGP_5$			$DGP_6$		
	$\delta$	n=50	100	200	n=50	100	200	n=50	100	200
$\alpha = 0.10$	0.0	0.186	0.152	0.100	0.164	0.074	0.080	0.182	0.146	0.098
	1.0	1.000	1.000	1.000	0.942	1.000	1.000	0.768	0.894	1.000
$\alpha = 0.05$	0.0	0.160	0.082	0.048	0.086	0.068	0.044	0.130	0.074	0.054
	1.0	0.998	1.000	1.000	0.910	1.000	1.000	0.532	0.826	1.000
$\alpha = 0.01$	0.0	0.046	0.038	0.008	0.036	0.068	0.008	0.082	0.056	0.010
	1.0	0.982	1.000	1.000	0.742	1.000	1.000	0.412	0.736	0.984
Case 2.1		$H_0$ : Overlapping varying coefficient model ( $\delta_2 = 1$ )								
		$DGP_4$			$DGP_5$			$DGP_6$		
	$\delta = \delta_1$	n=50	100	200	n=50	100	200	n=50	100	200
$\alpha = 0.10$	0.0	0.122	0.108	0.104	0.072	0.088	0.094	0.146	0.122	0.108
	1.0	1.000	1.000	1.000	0.906	1.000	1.000	0.708	1.000	1.000
$\alpha = 0.05$	0.0	0.058	0.044	0.042	0.038	0.044	0.064	0.076	0.066	0.050
	1.0	1.000	1.000	1.000	0.846	1.000	1.000	0.664	0.906	0.998
$\alpha = 0.01$	0.0	0.028	0.018	0.014	0.002	0.004	0.006	0.026	0.018	0.010
	1.0	0.998	1.000	1.000	0.774	1.000	1.000	0.404	0.786	0.898
Case 3		$H_0$ : Partially linear model								
		$DGP_4$			$DGP_5$			$DGP_6$		
	$\delta = \delta_1 = \delta_2$	n=50	100	200	n=50	100	200	n=50	100	200
$\alpha = 0.10$	0.0	0.112	0.138	0.104	0.162	0.124	0.096	0.144	0.120	0.114
	1.0	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$\alpha = 0.05$	0.0	0.068	0.072	0.054	0.082	0.064	0.060	0.070	0.062	0.054
	1.0	1.000	1.000	1.000	0.998	1.000	1.000	1.000	1.000	1.000
$\alpha = 0.01$	0.0	0.020	0.026	0.014	0.036	0.020	0.014	0.022	0.020	0.016
	1.0	1.000	1.000	1.000	0.994	1.000	1.000	0.954	1.000	1.000

*Note:* Empirical size and power are calculated based on 500 simulations with 299 bootstrap repetitions. The *rule of thumb* bandwidths have a scaling factor  $C = 1.0$ , and  $\alpha$  is the significant level.

which directly compares the  $L_2$  norm of  $\hat{m}(W) - \hat{m}_{add}(W)$ , the functional difference between the fully nonparametric and additive estimate, contains a bias from estimating a nonlinear additive function. Furthermore, the existing choice of optimal bandwidth for regression may not be appropriate for the additivity test.

Wang and Carriere (2011) propose a different test statistics  $\hat{T}_{wc}$ , which first performs a conditional smoothing on the residual from an additive estimation, then constructs a  $L_2$  norm of the smoothed residual.

$$\hat{T}_{wc} = \int_{\mathbb{R}^d} \hat{E}(y - \hat{m}_{add}(W)|W)^2 dF(W), \quad (1.6)$$

where  $\hat{E}(\cdot|W)$  represents the Nadaraya-Watson estimator, and  $\hat{m}_{add}(W) = \hat{\mu} + \sum_{j=1}^d \hat{m}_j(W_j)$ , with  $\hat{\mu} = \frac{1}{n} \sum_{i=1}^n y_i$  and  $\hat{m}_j(W_j)$  being the Marginal Integration estimate obtained as in Linton and Nielsen (1995). The test is shown to ameliorate the bias influence from estimating the additive components of the nonparametric regression, and requires less on the choice of the bandwidth. Specifically, an optimal bandwidth for regressions can be feasible when the dimension is higher than two.

The second additivity test by Sperlich et al. (2002) (STY hereafter) proposes a derivative-based  $\hat{T}_{sty}$  for interactions between two variables in an additive model. One can apply this test to any pair of different  $W_k$  and  $W_l$  for  $k \neq l$ , and  $k, l = 1 \dots, d$ , and this can be regarded as a test for separability/additivity in the regression model. For example, to test for additivity between  $W_k$  and  $W_l$ , we can use

$$\hat{T}_{sty} = \int_{\mathbb{R}^2} \left( \hat{m}^{(1,1)}(W_k, W_l) \right)^2 dF(W_k, W_l), \quad (1.7)$$

where  $\hat{m}^{(1,1)}(W_k, W_l)$  is an estimate for the cross derivative  $m^{(1,1)}(W_k, W_l) \equiv (\partial^2 m(W) / \partial W_k \partial W_l)$ . A vanishing  $\hat{T}_{sty}$  indicates that the cross derivative is close to zero everywhere, supporting additivity between  $W_k$  and  $W_l$ . In practice, for example,  $\hat{T}_{sty}$  can be implemented by calculating  $\hat{m}^{(1,1)}(W_k, W_l)$  from a special local quadratic estimator, estimating the directions of  $W_k$  and  $W_l$  through the local quadratic estimation, and the other directions with the local constant estimation.

We compare the performance of the three bootstrap tests  $\hat{T}^*$ ,  $\hat{T}_{wc}^*$ , and  $\hat{T}_{sty}^*$  with the following three DGPs adapted from Sperlich et al. (2002),

$$DGP_7 : y_i = 0.5 + 2x_i + 1.5 \sin(-1.5z_{1i}) + \delta_1 m_{12}(x_i, z_{1i}) + \epsilon_i$$

$$DGP_8 : y_i = 0.5 + 1.5 \sin(-1.5x_i) - z_{1i}^2 + \delta_1 m_{23}(x_i, z_{1i}) + \epsilon_i$$

$$DGP_9 : y_i = 0.5 + 2x_i - z_{1i}^2 + \delta_1 m_{13}(x_i, z_{1i}) + \epsilon_i$$

where  $x_i, z_{1i}$  and  $\epsilon_i$  are generated as in the first study. For a non-zero value of  $\delta_1$ , the presence of interaction functions  $m_{12}(x_i, z_{1i}) = x_i z_{1i}$ ,  $m_{23}(x_i, z_{1i}) = x_i e^{z_{1i}}$ , and  $m_{13}(x_i, z_{1i}) = x_i \sin(z_{1i})$  reflects the alternative non-additive structure. When  $\delta_1 = 0$ ,  $DGP_7 - DGP_9$  become purely additive. Recall that when  $d = 2$ , our test  $\hat{T}^*$  is capable of testing  $x$ - $z_1$  separability under the null that  $\omega = z_1$  is insignificant in  $g_x(W)$  in Case 1. Hence, we choose to construct our test based on Case 1.

To test the null  $H_0 : m(W) = m_{add}(W)$ , the bootstrap tests  $\hat{T}_{wc}^*$  and  $\hat{T}_{sty}^*$  require estimation of the additive functions, say,  $\hat{m}_j(W_j)$ , with  $j = 1, 2$ , through a Marginal Integration step that essentially averages a bivariate estimate of  $m(\cdot)$  over the other direction. Following their suggestions, we utilize two potentially different bandwidths, one for the direction of interest ( $h_1$ ) and the other for the other direction to be averaged over ( $h_2$ ). Sperlich et al. (2002) and Wang and Carriere (2011) both suggest using a double-bandwidth strategy that utilizes another set of bandwidths  $h$  for the testing step, potentially different from the bandwidths used in the estimation step. In contrast, our test  $\hat{T}^*$  only requires a single set of bandwidths  $h$ .

For a fair comparison, we implement the three test statistics with the same Gaussian kernel function and carefully choose the bandwidth sequences. Since the optimal bandwidths for regression estimation are compatible with our assumptions, we start with the *rule-of-thumb* bandwidth  $h_\xi = C\hat{\sigma}_\xi n^{-\frac{1}{6}}$ , where  $\xi = x$  or  $z_1$ . We have used  $C = 1$  in our first two studies, however, the test  $\hat{T}_{sty}^*$  requires a relatively large bandwidth (see Assumption A.6 in Sperlich et al. (2002)), so we choose  $C = 1.5$  and call the set of bandwidths selected as  $h$ .<sup>2</sup> Though the implementation of both  $\hat{T}_{sty}^*$  and  $\hat{T}_{wc}^*$  recommends different bandwidths, i.e.,  $h_1$  for the direction of interest in the additive function estimation and  $h_2$  for the other direction to be averaged over, the recommended  $h_1$ ,  $h_2$  and  $h$  in the simulation of Sperlich et al. (2002) do not change much in magnitude, thus we use  $h$  discussed above to implement  $\hat{T}_{sty}^*$ . In contrast, WC argue for the importance of the double bandwidth strategy and utilize  $h$ ,  $h_1$  and  $h_2$  of very different magnitudes in their simulation. To mimic their choice of bandwidths, we use  $h$  as above to implement their test, but for the estimation bandwidths, we set  $C = 0.5$  in  $h_\xi = C\hat{\sigma}_\xi n^{-\frac{1}{6}}$  and call the resulted bandwidths  $h_1$  for the direction of interest (undersmoothed), and set  $C = 1$  and call the resulted bandwidths  $h_2$  for the other direction to be averaged over (oversmoothed).<sup>3</sup> We choose the sample sizes  $n = (50, 100, 200)$ , perform 500 simulations with 299 bootstrap repetitions to evaluate the empirical size by setting  $\delta_1 = 0$ , and power with  $\delta_1 = (0.2, 0.4, \dots, 1.0)$ , respectively, and we set the significance level  $\alpha$  to be 5%.

Table 1.3 presents the size and power under  $DGP_7$ ,  $DGP_8$ , and  $DGP_9$  in Panel A, B, and C, respec-

<sup>2</sup>We observe from our simulation that an oversmoothing bandwidth  $h$  is critical for the power of  $\hat{T}_{sty}^*$ . For example, when  $C$  is less than 1.3, the power of  $\hat{T}_{sty}^*$  is small and not changing much with sample sizes, making the comparisons difficult. Thus, we adopt a larger constant  $C = 1.5$  for all three tests.

<sup>3</sup>For  $\hat{T}_{wc}^*$  implementation, we have also set  $C = (0.4, 1.6)$ ,  $C = (0.3, 1.7)$  for the constant in the bandwidth of  $(h_1, h_2)$ . The conclusion of the tests comparison in terms of size and power does not change qualitatively. Thus we do not report these results to save space.



Table 1.3: Additivity Test Comparison From A Bivariate Regression

Panel A			DGP <sub>7</sub>						
$\delta_1$	$\hat{T}^*$	n=50		$\hat{T}^*$	n=100		$\hat{T}^*$	n=200	
		$\hat{T}_{wc}^*$	$\hat{T}_{sty}^*$		$\hat{T}_{wc}^*$	$\hat{T}_{sty}^*$		$\hat{T}_{wc}^*$	$\hat{T}_{sty}^*$
0.0	0.076	0.092	0.048*	0.056*	0.102	0.084	0.048*	0.066	0.080
0.2	0.252	0.178	0.168	0.324	0.524	0.212	0.528	0.968	0.278
0.4	0.744	0.348	0.356	0.954	0.890	0.490	1.000	1.000	0.682
0.6	0.938	0.538	0.516	1.000	0.968	0.662	1.000	1.000	0.856
0.8	0.996	0.578	0.644	1.000	0.984	0.776	1.000	1.000	0.904
1.0	1.000	0.656	0.612	1.000	0.986	0.808	1.000	1.000	0.946

Panel B			DGP <sub>8</sub>						
$\delta_1$	$\hat{T}^*$	n=50		$\hat{T}^*$	n=100		$\hat{T}^*$	n=200	
		$\hat{T}_{wc}^*$	$\hat{T}_{sty}^*$		$\hat{T}_{wc}^*$	$\hat{T}_{sty}^*$		$\hat{T}_{wc}^*$	$\hat{T}_{sty}^*$
0.0	0.030	0.136	0.062*	0.040	0.146	0.054*	0.052	0.184	0.050*
0.2	0.186	0.328	0.284	0.406	0.648	0.456	0.810	0.994	0.748
0.4	0.820	0.448	0.526	0.996	0.920	0.696	1.000	1.000	0.868
0.6	0.990	0.526	0.634	1.000	0.964	0.708	1.000	1.000	0.922
0.8	1.000	0.574	0.586	1.000	0.974	0.764	1.000	1.000	0.912
1.0	1.000	0.608	0.610	1.000	0.992	0.782	1.000	1.000	0.962

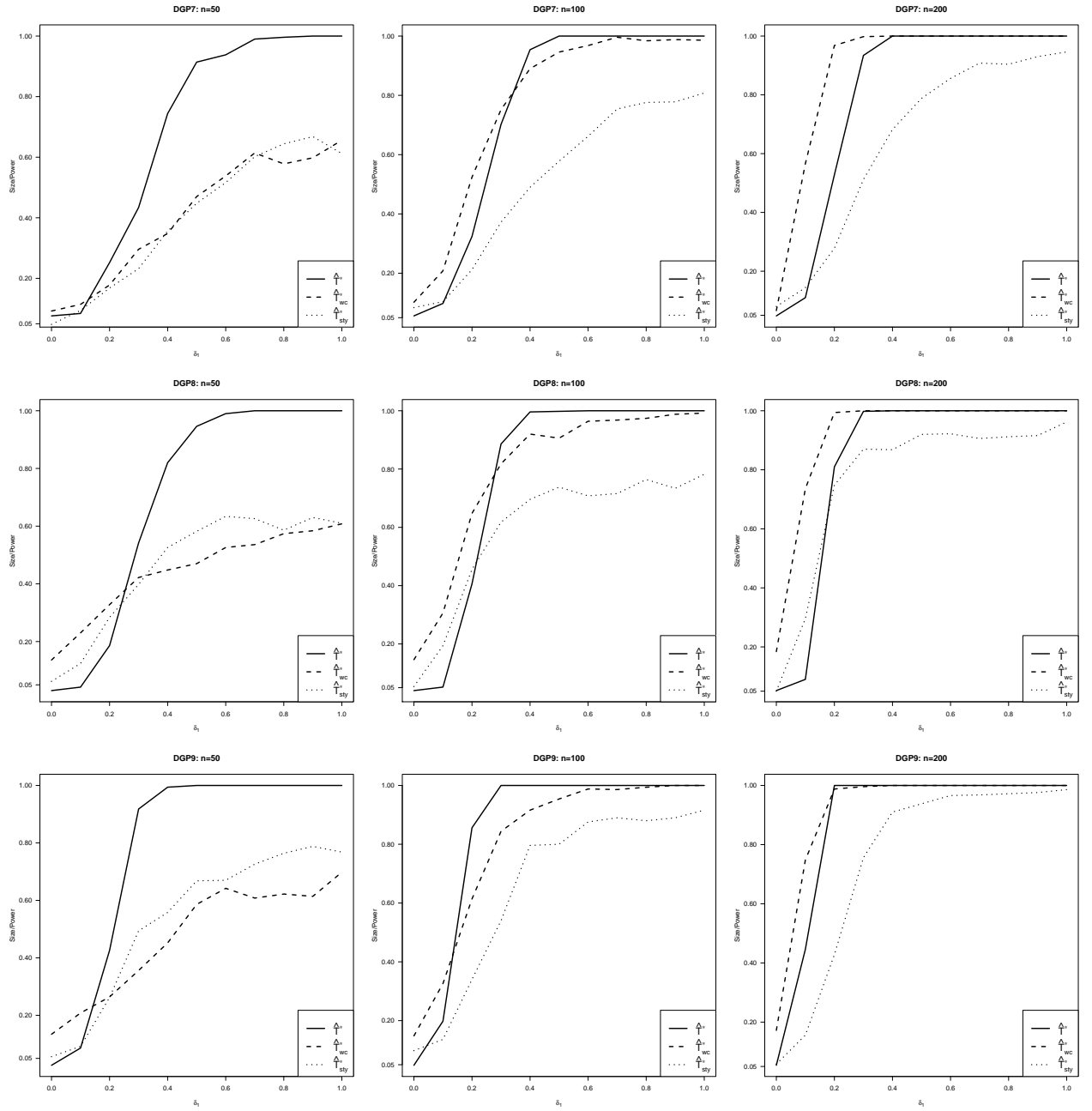
Panel C			DGP <sub>9</sub>						
$\delta_1$	$\hat{T}^*$	n=50		$\hat{T}^*$	n=100		$\hat{T}^*$	n=200	
		$\hat{T}_{wc}^*$	$\hat{T}_{sty}^*$		$\hat{T}_{wc}^*$	$\hat{T}_{sty}^*$		$\hat{T}_{wc}^*$	$\hat{T}_{sty}^*$
0.0	0.026	0.134	0.056*	0.048*	0.148	0.098	0.054*	0.172	0.058
0.2	0.426	0.264	0.262	0.856	0.614	0.340	1.000	0.988	0.428
0.4	0.994	0.452	0.558	1.000	0.916	0.796	1.000	1.000	0.910
0.6	1.000	0.642	0.670	1.000	0.988	0.876	1.000	1.000	0.966
0.8	1.000	0.622	0.764	1.000	0.994	0.880	1.000	1.000	0.972
1.0	1.000	0.698	0.768	1.000	1.000	0.916	1.000	1.000	0.986

*Note:* We report the empirical rejection frequency at 5% significant level from 500 simulations with 299 bootstrap repetitions. The test statistics  $\hat{T}_{wc}^*$  and  $\hat{T}_{sty}^*$  are short for, respectively, the additivity test statistic by Wang and Carriere (2011) and Sperlich et al. (2002).  $\hat{T}^*$  is the propose test statistics corresponding to Case 1 under the bivariate regression setting. All tests adopt Gaussian kernel with *rule-of-thumb* bandwidths with different scaling factor  $C$  discussed in Section 5.

tively. To facilitate the comparison of three tests' size, we put a  $\star$  on the upper right of the size closest to the nominal level.  $\hat{T}^*$  and  $\hat{T}_{sty}^*$  exhibit a fairly reasonable size performance, with  $\hat{T}^*$  being the closest to the nominal level in  $DGP_7$  and  $DGP_9$  for large samples, and  $\hat{T}_{sty}^*$  being the closest in  $DGP_8$ , followed closely by  $\hat{T}^*$ .  $\hat{T}_{wc}^*$  is generally over-sized, whose size approaches the target level only in  $DGP_7$ . Overall, our test captures the size fairly well relative to the other two tests.

The power of each test increases quickly, albeit at different rates as the sample size increases, or as  $\delta_1$  (deviation from the null) increases across  $DGP_{7-9}$ , indicating that all tests are consistent. To visualize the difference in power, we plot the power curve of the three tests in Figure 1.1, by setting  $\delta_1 = (0.1, 0.2, \dots, 1.0)$ . When  $n = 50$ ,  $\hat{T}^*$  performs the best in terms of a large power across all DGPs for any  $\delta_1 > 0.3$ . With  $n = 100$ , our test performs best when  $\delta_1 > 0.4$ , and the power of the three tests increases quickly, more so for  $\hat{T}_{wc}^*$ , followed by  $\hat{T}^*$ , and then by  $\hat{T}_{sty}^*$ . This observation is in fact expected.

Figure 1.1: Power Curve Comparison of Additivity Tests at 5% Rejection Region



When  $d = 2$  the convergence rate of  $\hat{T}_{wc}^*$  is  $nh$  (Theorem 3 in WC), faster than the rate of  $nh^3$  in our test  $\hat{T}^*$  and  $nh^5$  in  $\hat{T}_{sty}^*$  (Theorem 9 in STY). With  $n = 200$ , our test and  $\hat{T}_{wc}^*$  all have power one when the model only moderately deviates from its null ( $\delta_1 > 0.4$  in  $DGP_{7-8}$  and  $\delta_1 > 0.2$  in  $DGP_9$ ), while the power of  $\hat{T}_{sty}^*$  is increasing but still distinguishably less than one.  $\hat{T}_{sty}^*$  exhibits the largest power with a smaller  $\delta_1$ . In all, our test  $\hat{T}^*$  exhibits a reasonable performance of size and power, presenting a viable alternative to the other two additivity tests considered in the simulation study.

## 1.6 Empirical Application

Revealing the salient semiparametric structure of the regression model can be highly useful. First, evidence of a correct semiparametric structure can be usefully incorporated into the model, improving the estimation efficiency and alleviating the *curse of dimensionality*. Second, the revealed structure can further be used to test for the validity of the underlying economic hypothesis. Tests for a specific semiparametric/parametric structure either fail to reject or reject the null hypothesis, and in the case of rejecting the null, the practitioner is left with only the knowledge of an incorrect null structure, but no hint regarding where to further search for the feature of the structure, if there is any. One interesting aspect of our test is that repeated use of our test can further shed light on the structure of the regression model, since the information on the significant or insignificant variables in the gradient function can be translated into information regarding the structure of the regression function, as we show in the simulation section with three popular semiparametric models. We illustrate the empirical applicability of our proposed test statistic in this section.

We utilize the dataset consisting of 569 Belgian firms in 1996 from Verbeek (2008), regarding the estimation of a labor demand function. The dataset is available in **Labour** under R package **Ecdat**. Economic theories suggest that the labor demand  $L$  is a function of the capital stock ( $K$ ), output ( $Y$ ), and real labor's wage ( $w$ ), however, there is no clear guidance regarding the specific structure or the functional form of the labor demand function. For example, modeling the labor demand in a simple linear fashion without interactive effects among its key determinants is likely to be misspecified. We implement our test to provide statistical evidence on the potential nonlinearity and interactive effects, illustrating to empirical practitioners a plausible underlying structure of the labor demand.

We start by assuming that the labor demand function is a fully unknown nonparametric function with an additive error, i.e.,

$$L_i = m(K_i, Y_i, w_i) + \epsilon_i, \quad (1.8)$$

where  $L$  ( $\ln$  labor) is the natural log of total number of workers, depending on **capital** ( $K$ , total fixed assets

Table 1.4: Descriptive Statistics of the Dataset

Variable	Index	n	Mean	SD	Min	Max
ln labor	$L$	481	4.355	1.048	0.000	7.185
capital	$K$	481	3.982	5.472	0.002	42.235
output	$Y$	481	6.571	7.011	0.026	48.452
wage	$w$	481	34.587	7.146	11.734	49.892

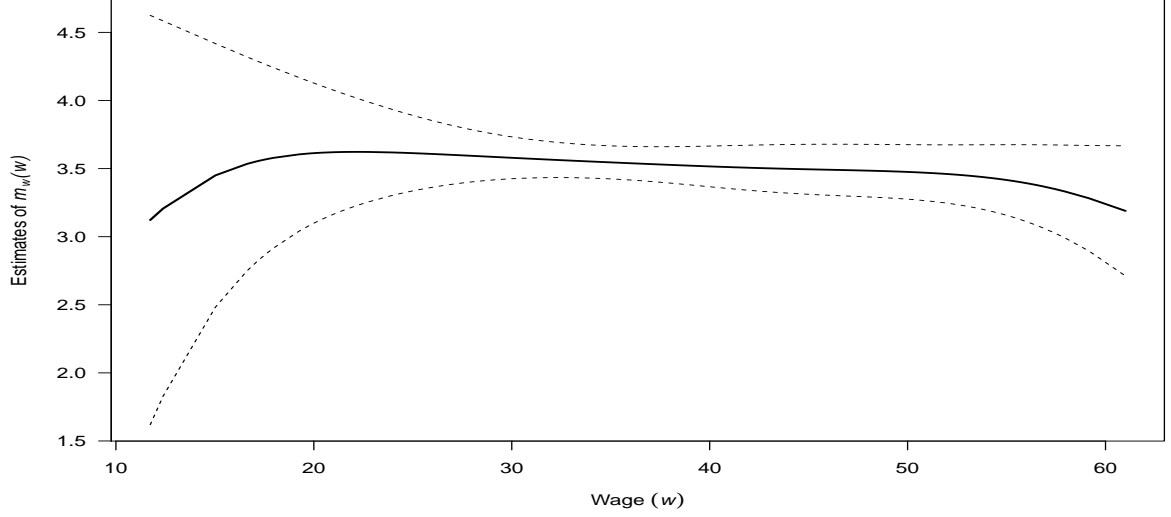
Table 1.5: Empirical Results From the Dataset

<i>Panel A</i>		Gradient of $K$				
Test	$\omega$	$\omega^c$	$\hat{T}^*$	p-value	$\hat{T}^*$ (trim)	p-value (trim)
1	$K, Y, w$	-	0.0043	0.0725	0.0039	0.0875
2	$K$	$Y, w$	0.0014	0.1200	0.0008	0.1825
3	$Y$	$K, w$	0.0235	0.0000	0.0212	0.0002
4	$w$	$K, Y$	0.0017	0.2200	0.0019	0.2452
Inferred model structure		$L = Km_1(Y) + m_2(Y, w) + u$				
<i>Panel B</i>		Gradient of $Y$				
Test	$\omega$	$\omega^c$	$\hat{T}^*$	p-value	$\hat{T}^*$ (trim)	p-value (trim)
5	$K, Y, w$	-	0.0282	0.0000	0.0258	0.0000
6	$Y$	$K, w$	0.0235	0.0000	0.0203	0.0000
7	$w$	$K, Y$	0.0033	0.1700	0.0024	0.1937
Inferred model structure		$L = Km_1(Y) + m_Y(Y) + m_w(w) + u$				
<i>Panel C</i>		Gradient of $w$				
Test	$\omega$	$\omega^c$	$\hat{T}^*$	p-value	$\hat{T}^*$ (trim)	p-value (trim)
8	$K, Y, w$	-	0.0007	0.2475	0.0005	0.2898
Inferred model structure		$L = Km_1(Y) + m_Y(Y) + w\beta + u$				

in million euro), **output** ( $Y$ , value added in million euro), and **wage** ( $w$ , total wage costs divided by number of workers in 1000 euro). To mitigate the impact of extreme outliers, we trim out firms falling in the upper ten percentiles of each variable. Table 1.4 provides the descriptive statistics for the 481 observations in the dataset. Before proceeding, a verification that all regressors in (1.8) are significant in  $m(\cdot)$  is needed, as a necessary condition to proceed with our test. We implement a nonparametric  $R^2$  test for omitted variables by Yao and Ullah (2013), to check whether all regressors in (1.8) are significant in  $m(\cdot)$ . Their standardized bootstrap test  $\hat{T}_{nG}^*$  with a *rule-of-thumb* bandwidth and 399 bootstrap repetitions gives a p-value of 0.0000, clearly rejecting the null of insignificant  $K, Y$  and  $w$  in  $m(\cdot)$ .<sup>4</sup>

We start the investigation by testing the significance of  $\omega$  in the gradient function  $g_K(W)$ , with  $W = [K, Y, w]'$  and the results are summarized in Panel A of Table 1.5. We begin by specifying  $\omega = W$  in the null of Test 1, which is equivalent to testing the structure that  $K$  is linear and not interactive with  $Y$  and  $w$ . Result of Test 1 from Panel A of Table 1.5 show that the null is rejected at 10% significance level,

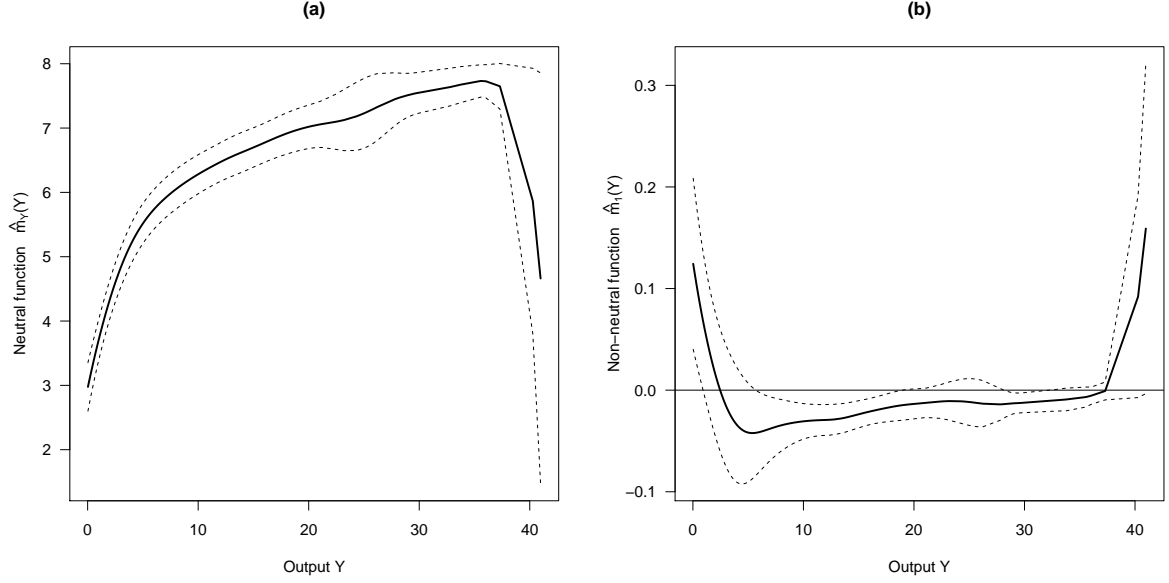
<sup>4</sup>The results are omitted here for brevity and are available upon the request from the second author.

Figure 1.2: Plot of  $m_w(w)$  estimates against  $w$ 

suggesting that a partially linear model for the labor demand as  $m(K, Y, w) = K\beta + m(Y, w)$  is not likely to be true. Intuitively, the rejection of the null in Test 1 is likely due to the presence of nonlinearity of  $K$ , interaction effects between  $K$  and the other variables ( $Y, w$ ), or both. Hence, we proceed to test the nonlinearity of  $K$  in Test 2 with the null being  $\omega = K$ , the interaction effect between  $K$  and  $Y$  in Test 3 with  $\omega = Y$ , and the interaction effect between  $K$  and  $w$  in Test 4 with  $\omega = w$ . Panel A of Table 1.5 indicates that we fail to reject the null in Tests 2 and 4 but reject the null in Test 3 at 1% significant level. Thus, we conclude that  $K$  (capital) enters the labor demand function linearly, whose coefficient varies significantly with  $Y$  (output) but not with  $w$  (wage). To summarize, the results in Panel A suggest that the structure of the labor demand function may be characterized as  $m(K, Y, w) = Km_1(Y) + m_2(Y, w)$ , a varying coefficient model but with different smoothing variables.

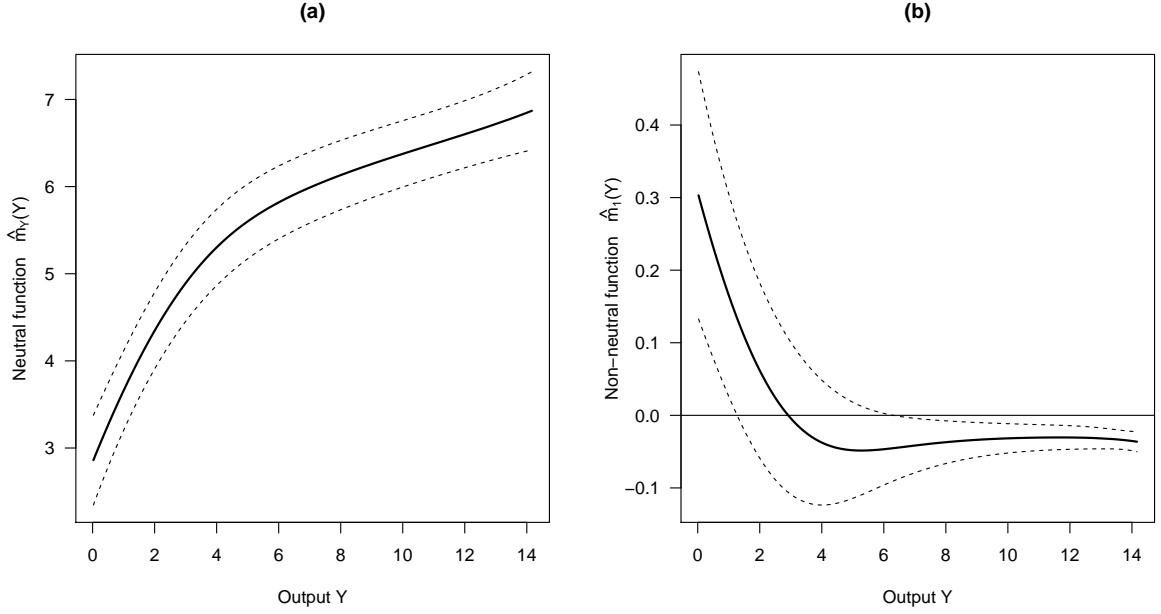
We further explore the structure of  $m_2(Y, w)$ , which is shown from the tests in Panel A to be additively separable from  $Km_1(Y)$ , and more detailed structure can surely improve the estimation efficiency. For illustration purposes, we explore the structure of  $m_2(Y, w)$  by testing the significance of  $\omega$  in the gradient function  $g_Y(W)$ . The p-values in Panel B of Table 1.5 indicate that the effect of  $Y$  is significantly nonlinear (in Tests 5 and 6) but not interactive with  $w$  (in Test 7). Thus, given the tests in Panel A, we expect that  $m_2(\cdot)$  is additively separable, i.e.,  $m_2(Y, w) = m_Y(Y) + m_w(w)$ . Hence, our labor demand function structure is pinned down to  $m(K, Y, w) = Km_1(Y) + m_Y(Y) + m_w(w)$ , a varying coefficient additive model.

Finally, one may conjecture that the effect of  $w$  is constant, (i.e.,  $m_w(w)$  is linear), since a linear specification is an empirically popular choice. Based on the revealed varying coefficient additive model

Figure 1.3: Plots of  $\hat{m}_Y(Y)$  and  $\hat{m}_1(Y)$  in the Partially Linear Varying Coefficient Model

structure in Panel B, i.e.,  $m(K, Y, w) = Km_1(Y) + m_Y(Y) + m_w(w)$ , we consistently estimate  $m_w(\cdot)$  through a polynomial B-Spline estimator, with the number of evenly-spaced interior knots chosen with the generalized cross-validation (see Li and Racine (2007)). We plot the estimates  $\hat{m}_w(w)$  against  $w$  with its 95% bootstrap CI superimposed in Figure 1.2, where  $\hat{m}_w(w)$  is demonstrated to be a downward sloping linear line except for the boundary area. Therefore, we further test the null hypothesis that  $\omega = W$  is insignificant in  $g_w(W)$ , corresponding to the structure that  $w$  exhibits a constant effect on the labor demand. The result of Test 8 in Panel C of Table 1.5 supports this conjecture, showing that  $w$  enters the model linearly, with no interactions with either  $K$  or  $Y$ , consistent with results in Tests 4 and 7. To conclude, the series of our tests reveal that the structure of the labor demand in the dataset is likely to be  $m(K, Y, w) = Km_1(Y) + m_Y(Y) + w\beta$ , a partially linear varying coefficient model (PLVCM).

We estimate the inferred PLVCM structure for the labor demand, i.e.,  $L_i = K_i m_1(Y_i) + m_Y(Y_i) + w_i \beta + \epsilon_i$ , to illustrate the interaction between  $Y$  and  $K$ , the nonlinearity of  $Y$ , and the constant effect of  $w$ . We follow Fan and Huang (2005) to estimate the neutral function  $m_Y(Y)$ , non-neutral function  $m_1(Y)$ , and constant partial effect  $\beta$  through the profile least-square estimator. The estimated neutral and non-neutral functions  $\hat{m}_Y(Y)$  and  $\hat{m}_1(Y)$  are plotted in Panel (a) and (b) of Figure 1.3, respectively, with the cross validation least square (CVLS) bandwidth implemented. The estimated neutral function of  $Y$  in panel (a) is clearly nonlinear and concave, suggesting a positive yet diminishing effect of the output on the labor demand. The coefficient function of  $K$  on  $L$  in panel (b) evidently depends on  $Y$  in a nonlinear fashion, which decreases to a negative level for relatively low level of outputs but steadily

Figure 1.4: Plots of  $\hat{m}_Y(Y)$  and  $\hat{m}_1(Y)$  in the Partially Linear Varying Coefficient Model

increases, and eventually turns positive as the output increases to a high level. This may be intuitively understood, as higher outputs produced may alter the labor-capital elasticity of substitution, making capital to be eventually complementary with labor in the production process. Finally, the parameter estimate is  $\hat{\beta} = -0.0278$  with a standard error of 0.0032. It indicates that  $w$  (**wage**) enters the model linearly and significantly, and one unit increase in the wage decreases the labor demand by about 2.78%.

We further check the sensitiveness of our testing results to the presence of large firms. We trim out firms with the upper 10% output, and re-perform the testing procedure as above. The results are reported on the last two columns of Table 1.5. We observe that our implied model structure remain unchanged, which continues to be a PLVCM. We plot the PLVCM estimates in Figure 1.4, which again reveals the nonlinearity of both neutral and non-neutral function of output in labor demand. The parametric estimate is  $\hat{\beta} = -0.029$  with a standard error of 0.0037, showing that wage enters the model linearity as indicated in our previous finding.

## 1.7 Conclusion

Given a set of significant regressors in conditional mean function, we propose a variation based test for significant variables in the gradient function. We construct the test by estimating the gradient with a local polynomial estimator, obtain the asymptotic null distribution of the test, and demonstrate that the test is consistent. Based on the analysis of the asymptotic null distribution, we propose a wild bootstrap test,

which is easy to implement. We demonstrate the validity of the bootstrap as its asymptotic distribution is the same as the asymptotic null, no matter whether the null hypothesis is valid or not.

The Monte Carlo studies illustrate encouraging finite sample properties of our test, with the size being close to the nominal level and the power approaching one quickly as the sample size increases. The test can be used to infer several well-known semiparametric structures, including the additive model, the varying coefficient model, and the partially linear model. Through an empirical application on estimating the labor demand, we illustrate that repeated use of our test can help empirical practitioners to infer the refined structure for the underlying regression, offering promises for the future work.

## Appendix

**Lemma 1.** *Uniformly for all  $W_i \in G$ , a compact subset of  $\mathbb{R}^d$ ,*

$$\hat{g}_x(W_i) - g_x(W_i) = \frac{1}{nh^{d+1}f(W_i)} \sum_{j=1}^n SK\left(\frac{W_j - W_i}{h}\right) \left( \sum_{|\mathbf{k}|=p+1} \frac{h^{p+1}}{\mathbf{k}!} (D^{\mathbf{k}}m)(W_i + \lambda(W_j - W_i)) \left(\frac{W_j - W_i}{h}\right)^{\mathbf{k}} + \epsilon_j \right) (1 + o_p(1)).$$

*Proof.* From the first order condition in equation (1.2), let's arrange the  $N_{|\mathbf{j}|}$  values of  $t_{n,\mathbf{j}}(W)$  in a column vector  $\tau_{n,|\mathbf{j}|}(W)$ , with the  $k$ -th element being  $(\tau_{n,|\mathbf{j}|})_k(W) = t_{n,G_{|\mathbf{j}|}(k)}(W)$ . For  $N = \sum_{i=0}^p N_i$ , we define

$$\tau_n(W) = \begin{bmatrix} \tau_{n,0}(W) \\ \tau_{n,1}(W) \\ \vdots \\ \tau_{n,p}(W) \end{bmatrix}. \text{ We arrange the distinct values of } h^{|\mathbf{k}|} \hat{\alpha}_{\mathbf{k}}(W) \text{ for } 0 \leq |\mathbf{k}| \leq p \text{ as a } N \times 1 \text{ column}$$

$$\text{vector } \hat{\alpha}_n(W) = \begin{bmatrix} \hat{\alpha}_{n,0}(W) \\ \hat{\alpha}_{n,1}(W) \\ \vdots \\ \hat{\alpha}_{n,p}(W) \end{bmatrix}, \text{ where } (\hat{\alpha}_{n,|\mathbf{j}|})_k(W) = h^{|\mathbf{j}|} \hat{\alpha}_{n,G_{|\mathbf{j}|}(k)}(W). \text{ For the true values, } \alpha(W) =$$

$$\begin{bmatrix} \alpha_0(W) \\ \alpha_1(W) \\ \vdots \\ \alpha_p(W) \end{bmatrix}, \text{ where } (\alpha_{|\mathbf{j}|})_k(W) = h^{|\mathbf{j}|} \alpha_{G_{|\mathbf{j}|}(k)}(W). \text{ Since we arrange } W = (x, z')' \text{ with } x \text{ being the first}$$

element,  $g_x(W) = a_{G_1(d)}(W)$ . Next we arrange the possible values of  $S_{n,\mathbf{j}+\mathbf{k}}(W)$  by a matrix  $S_{n,|\mathbf{j}|,|\mathbf{k}|}(W)$  in a lexicographical order with the  $(l, m)$ -th element being  $(S_{n,|\mathbf{j}|,|\mathbf{k}|})_{l,m}(W) = S_{n,G_{|\mathbf{j}|}(l)+G_{|\mathbf{k}|}(m)}(W)$ . So



$S_{n,|\mathbf{j}|,|\mathbf{k}|}(W)$  has dimension  $N_{|\mathbf{j}|} \times N_{|\mathbf{k}|}$ . Define  $S_n(W) = \begin{bmatrix} S_{n,0,0}(W) & S_{n,0,1}(W) & \cdots & S_{n,0,p}(W) \\ S_{n,1,0}(W) & S_{n,1,1}(W) & \cdots & S_{n,1,p}(W) \\ \vdots & \vdots & \cdots & \vdots \\ S_{n,p,0}(W) & S_{n,p,1}(W) & \cdots & S_{n,p,p}(W) \end{bmatrix}$ .

The set of equations in (1.2) are  $\tau_n(W) = S_n(W)\hat{\alpha}_n(W)$ . Assuming that  $S_n(W)$  is positive definite, the solution is expressed as  $\hat{\alpha}_n(W) = S_n^{-1}(W)\tau_n(W)$ . Specifically, let  $e_{N,1+d}$  be a  $N \times 1$  vector of zeros, except one at its  $(1+d)th$  position. Then

$$h\hat{g}_x(W) = e'_{N,1+d}S_n^{-1}(W)\tau_n(W).$$

We note that  $S_{n,G_{|\mathbf{j}|}(l)+G_{|\mathbf{k}|}(m)}(W) = \frac{1}{nh^d} \sum_{i=1}^n K\left(\frac{W_i-w}{h}\right) \left(\frac{W_i-W}{h}\right)^{G_{|\mathbf{j}|}(l)+G_{|\mathbf{k}|}(m)}$ . By assumption A3(2), for  $W^* = \lambda W_i + (1-\lambda)W$  with  $\lambda \in (0,1)$ ,

$$y_i = m(W_i) + \epsilon_i = \sum_{0 \leq |\mathbf{k}| \leq p} \frac{1}{\mathbf{k}!} (D^{\mathbf{k}}m)(W)(W_i - W)^{\mathbf{k}} + \sum_{|\mathbf{k}|=p+1} \frac{1}{\mathbf{k}!} (D^{\mathbf{k}}m)(W^*)(W_i - W)^{\mathbf{k}} + \epsilon_i.$$

$$\begin{aligned} t_{n,\mathbf{j}}(W) &= \frac{1}{nh^d} \sum_{i=1}^n K\left(\frac{W_i-w}{h}\right) \left(\frac{W_i-W}{h}\right)^{\mathbf{j}} \left[ \sum_{0 \leq |\mathbf{k}| \leq p} \frac{1}{\mathbf{k}!} (D^{\mathbf{k}}m)(W)(W_i - W)^{\mathbf{k}} + \sum_{|\mathbf{k}|=p+1} \frac{1}{\mathbf{k}!} (D^{\mathbf{k}}m)(W^*)(W_i - W)^{\mathbf{k}} + \epsilon_i \right] \\ &= \sum_{0 \leq |\mathbf{k}| \leq p} a_{\mathbf{k}} h^{|\mathbf{k}|} S_{n,\mathbf{j}+\mathbf{k}}(W) + \frac{1}{nh^d} \sum_{i=1}^n K\left(\frac{W_i-w}{h}\right) \left(\frac{W_i-W}{h}\right)^{\mathbf{j}} \left[ \sum_{|\mathbf{k}|=p+1} \frac{1}{\mathbf{k}!} (D^{\mathbf{k}}m)(W^*)(W_i - W)^{\mathbf{k}} + \epsilon_i \right] \\ &= \sum_{0 \leq |\mathbf{k}| \leq p} \hat{a}_{\mathbf{k}} h^{|\mathbf{k}|} S_{n,\mathbf{j}+\mathbf{k}}(W), \end{aligned}$$

where the last equality is from equation (1.2). Thus,

$$\sum_{0 \leq |\mathbf{k}| \leq p} h^{|\mathbf{k}|} (\hat{a}_{\mathbf{k}}(W) - a_{\mathbf{k}}(W)) S_{n,\mathbf{j}+\mathbf{k}}(W) = \frac{1}{nh^d} \sum_{i=1}^n K\left(\frac{W_i-w}{h}\right) \left(\frac{W_i-W}{h}\right)^{\mathbf{j}} \left[ \sum_{|\mathbf{k}|=p+1} \frac{1}{\mathbf{k}!} (D^{\mathbf{k}}m)(W^*)(W_i - W)^{\mathbf{k}} + \epsilon_i \right]. \quad (1.9)$$

Then we let  $\kappa_{i,|\mathbf{j}|}\left(\frac{W_i-W}{h}\right)$  be an  $N_{|\mathbf{j}|}$  dimensional subvector whose  $k-th$  element is  $[\kappa_{i,|\mathbf{j}|}\left(\frac{W_i-W}{h}\right)]_k =$

$$K\left(\frac{W_i-w}{h}\right) \left(\frac{W_i-W}{h}\right)^{G_{|\mathbf{j}|}(k)}. \text{ Furthermore, } \kappa_i\left(\frac{W_i-W}{h}\right) = \begin{bmatrix} \kappa_{i,0}\left(\frac{W_i-W}{h}\right) \\ \kappa_{i,1}\left(\frac{W_i-W}{h}\right) \\ \vdots \\ \kappa_{i,p}\left(\frac{W_i-W}{h}\right) \end{bmatrix}.$$

Thus we express the equations in (1.9) in a matrix format

$$S_n(W)(\hat{\alpha}_n(W) - \alpha(W)) = \frac{1}{nh^d} \sum_{i=1}^n \kappa_i\left(\frac{W_i-W}{h}\right) \left[ \sum_{|\mathbf{k}|=p+1} \frac{1}{\mathbf{k}!} (D^{\mathbf{k}}m)(W^*)(W_i - W)^{\mathbf{k}} + \epsilon_i \right].$$

So we have

$$h(\hat{g}_x(W) - g_x(W)) = e'_{N,1+d} S_n^{-1}(W) \frac{1}{nh^d} \sum_{i=1}^n \kappa_i \left( \frac{W_i - W}{h} \right) \left[ \sum_{|\mathbf{k}|=p+1} \frac{1}{\mathbf{k}!} (D^{\mathbf{k}} m)(W^*) (W_i - W)^{\mathbf{k}} + \epsilon_i \right]. \quad (1.10)$$

We note that the element of  $S$  are simply multivariate moment of  $K(\cdot)$ , corresponding to the element of  $S_n$ . Consider a typical element of  $S_n - f(W)S$  as

$$[S_{n,i,j} - f(W)S_{i,j}]_{l,m} = S_{n,G_i(l)+G_j(m)} - ES_{n,G_i(l)+G_j(m)} + ES_{n,G_i(l)+G_j(m)} - \mu_{k,G_i(l)+G_j(m)} f(W) = I_1 + I_2.$$

$I_1 = \frac{1}{nh^d} \sum_{i=1}^n K \left( \frac{W_i - W}{h} \right) \left( \frac{W_i - W}{h} \right)^{G_i(l)+G_j(m)} - E \frac{1}{nh^d} \sum_{i=1}^n K \left( \frac{W_i - W}{h} \right) \left( \frac{W_i - W}{h} \right)^{G_i(l)+G_j(m)}$ . With A1(2), A2 and A4(1), we easily obtain  $\sup_{W \in G} |I_1| = O_p((\frac{nh^d}{\ln n})^{-\frac{1}{2}})$  (see Lemma 1 in Martins-Filho et al. (2018)).

$I_2 = \int K(\Psi) \Psi^{G_i(l)+G_j(m)} (f(W + h\Psi) - f(W)) d\Psi = h \int K(\Psi) \Psi^{G_i(l)+G_j(m)} \Psi' f^{(1)}(W^*) d\Psi = O(h)$  uniformly  $\forall W \in G$ , by A1(2) and A2(2).

So  $\sup_{W \in G} \|[S_{n,i,j} - f(W)S_{i,j}]_{l,m}\| = O_p(h + (\frac{nh^d}{\ln n})^{-\frac{1}{2}}) = o_p(1)$ . Furthermore, given A5 and A1(2), we obtain  $\sup_{W \in G} \|S_n - Sf(W)\| = o_p(1)$ , where  $\|\cdot\|$  refers to the Euclidean norm. Since  $S$  is positive definite in A5, the smallest eigenvalue of  $S$  is greater than zero, then we have  $\sup_{W \in G} \|S_n^{-1} - \frac{1}{f(W)} S^{-1}\| = o_p(1)$ .

From equation (1.9), we have

$$h(\hat{g}_x(W) - g_x(W)) = \frac{1}{f(W)} [S^{-1}]_{(1+d),\cdot} \frac{1}{nh^d} \sum_{i=1}^n \kappa_i \left( \frac{W_i - W}{h} \right) \left[ \sum_{|\mathbf{k}|=p+1} \frac{1}{\mathbf{k}!} (D^{\mathbf{k}} m)(W^*) (W_i - W)^{\mathbf{k}} + \epsilon_i \right] (1 + o_p(1)), \quad (1.11)$$

where  $[S^{-1}]_{(1+d),\cdot}$  refers to the  $(1+d)th$  row of  $S^{-1}$ . With the definition of  $SK(\cdot)$  in equation (1.5), we obtain the claimed result.  $\square$

In our proof, we have made repeated use of the following Lemma 2, which is the same as Theorem 1 in Yao and Martins-Filho (2015). Let  $\{Q_i\}_{i=1}^n$  be a sequence of independent and identically distributed (IID) random variables and  $\phi_n(Q_1, \dots, Q_k)$  be a symmetric function with  $k < n$ . We call  $\phi_n(Q_1, \dots, Q_k)$  a kernel function that depends on  $n$  and a U-statistic  $u_n$  of degree  $k$  is defined as

$$u_n = \binom{n}{k}^{-1} \sum_{(n,k)} \phi_n(Q_{i_1}, \dots, Q_{i_k}), \quad (1.12)$$

where  $\sum_{(n,k)}$  denotes the sum over all subsets  $1 \leq i_1 < i_2 < \dots < i_k \leq n$  of  $\{1, 2, \dots, n\}$ . Now, let  $\phi_{cn}(q_1, \dots, q_c) = E(\phi_n(Q_1, \dots, Q_c, Q_{c+1}, \dots, Q_k) | Q_1 = q_1, Q_2 = q_2, \dots, Q_c = q_c)$ ,  $\sigma_{cn}^2 = Var(\phi_{cn}(Q_1, \dots,$

$Q_c)$  and  $\theta_n = E(\phi_n(Q_1, \dots, Q_k))$ . In addition, recursively define  $h_n^{(1)}(q_1) = \phi_{1n}(q_1) - \theta_n, \dots, h_n^{(c)}(q_1, \dots, q_c) = \phi_{cn}(q_1, \dots, q_c) - \sum_{j=1}^{c-1} \sum_{(c,j)} h_n^{(j)}(q_{i_1}, \dots, q_{i_j}) - \theta_n$  for  $c = 2, \dots, k$ , where the sum  $\sum_{(c,j)}$  is over all subsets  $1 \leq i_1 < \dots < i_j \leq c$  of  $\{1, \dots, c\}$ . By Hoeffding's H-decomposition we have

$$u_n = \theta_n + \binom{n}{k}^{-1} \sum_{j=1}^k \binom{n-j}{k-j} \sum_{(n,j)} h_n^{(j)}(Q_{v_1}, \dots, Q_{v_j}) = \theta_n + \sum_{j=1}^k \binom{k}{j} H_n^{(j)}(Q_{v_1}, \dots, Q_{v_j}),$$

where  $H_n^{(j)}(Q_{v_1}, \dots, Q_{v_j}) = \binom{n}{j}^{-1} \sum_{(n,j)} h_n^{(j)}(Q_{v_1}, \dots, Q_{v_j})$ . Since  $u_n$  can be written as a finite sum of  $H_n^{(j)}$ , its magnitude can be determined by studying  $H_n^{(j)}$ . The following result shows that the magnitude of  $H_n^{(j)}$  is determined by  $n$  and the leading variance  $\sigma_{jn}^2$  defined above.

**Lemma 2.** *Let  $\{Q_i\}_{i=1}^n$  be an IID sequence and  $u_n$  be defined as in equation (1.12) such that*

$$u_n = \theta_n + \sum_{j=1}^k \binom{k}{j} H_n^{(j)}(Q_{v_1}, \dots, Q_{v_j}).$$

Then,

- (a)  $\text{Var}(H_n^{(j)}) = O\left(n^{-j} \sum_{c=1}^j \sigma_{cn}^2\right) = O(n^{-j} \sigma_{jn}^2)$  and  $H_n^{(j)} = O_p\left((n^{-j} \sigma_{jn}^2)^{\frac{1}{2}}\right)$ ;
- (b) for  $1 \leq c \leq c' \leq k$ , we have  $\frac{\sigma_{cn}^2}{c} \leq \frac{\sigma_{c'n}^2}{c'}$ .

**Theorem 1.**

*Proof.* Following the arguments after Theorem 1, we show below that

$$(1) \quad nh^{2+\frac{d}{2}}(T_{21} - \frac{1}{nh^{2+d}}B_1) \xrightarrow{d} N(0, \Omega).$$

$$(2) \quad T_{23} = -\frac{2}{nh^{2+d-d_1}}B_3(1 + o_p(1)) + o_p((nh^{2+\frac{d}{2}})^{-1}).$$

$$(3) \quad T_{22} = \frac{1}{nh^{2+d-d_1}}B_2(1 + o_p(1)) + o_p((nh^{2+\frac{d}{2}})^{-1}).$$

(1)-(3) give the claimed results.

(1) As stated before,  $T_{21} = T_{211} + \dots + T_{214}$ .

$$(a) \quad T_{211} = \frac{1}{n^3} \sum_{i=1}^n \sum_{j=1}^n \sum_{t=1}^n \frac{1}{f(W_i)^2 h^{2d+2}} SK\left(\frac{W_i - W_i}{h}\right) SK\left(\frac{W_t - W_i}{h}\right) \epsilon_j \epsilon_t = \frac{1}{n^3} \sum_{i=1}^n \sum_{j=1}^n \sum_{t=1}^n \psi_{nijt}, \text{ and notice } \psi_{nijt}$$

is symmetric in indices  $j$  and  $t$ .

(i) Consider the case that  $i \neq j \neq t$ . Let  $\phi_{nijt} = \psi_{nijt} + \psi_{njit} + \psi_{ntji}$ , and it is symmetric in all of its indices, then

$$\begin{aligned}
T_{211} &= \frac{1}{3n^3} \sum_{i=1}^n \sum_{j=1}^n \sum_{t=1}^n \phi_{nijt} = \frac{2}{n^3} \sum_{i=1}^n \sum_{j=1}^n \sum_{t=1}^n \phi_{nijt} = \frac{1}{3} [O(n^{-4}) + \binom{n}{3}^{-1} \sum_{i=1}^n \sum_{j=1}^n \sum_{t=1}^n \phi_{nijt} \\
&= \frac{1}{3} \binom{n}{3}^{-1} \sum_{i=1}^n \sum_{j=1}^n \sum_{t=1}^n \phi_{nijt} (1 + o_p(1))
\end{aligned}$$

Note that it is a U-statistic. We apply the H-decomposition result for U-statistic with sample size dependent kernel in Lemma 2, for  $\phi_n(Q_i, Q_j, Q_t) \equiv \phi_{nijt}$ , to obtain

$$\begin{aligned}
T_{211} &= \frac{1}{3} [\theta_n + \sum_{j=1}^3 \binom{3}{j} H_n^{(j)}(Q_{v_1}, \dots, Q_{v_j})] (1 + o_p(1)), \text{ where} \\
H_n^{(j)}(Q_{v_1}, \dots, Q_{v_j}) &= \binom{n}{j}^{-1} \sum_{(n,j)} h_n^{(j)}(Q_{v_1}, \dots, Q_{v_j}), \quad Q_i = (W_i', \epsilon_i)'. \text{ Here, } \theta_n = E(\phi_{nijt}) = 0. \\
H_n^{(1)}(Q_{v_1}) &= \binom{n}{1}^{-1} \sum_{i=1}^n h_n^{(1)}(Q_i) = \frac{1}{n} \sum_{i=1}^n [E(\phi_{nijt}|Q_i) - \theta_n] = 0, \text{ since } E(\epsilon_i|W_i) = 0. \\
\frac{1}{3} \binom{3}{2} H_n^{(2)}(Q_{v_1}, Q_{v_2}) &= \binom{n}{2}^{-1} \sum_{i=1}^n \sum_{j=1}^n h_n^{(2)}(Q_i, Q_j) \\
&= \binom{n}{2}^{-1} \sum_{i=1}^n \sum_{j=1}^n [E(\phi_{nijt}|Q_i, Q_j) - h_n^{(1)}(Q_i) - h_n^{(1)}(Q_j) - \theta_n] \\
&= \binom{n}{2}^{-1} \sum_{i=1}^n \sum_{j=1}^n E(\psi_{nijt} + \psi_{njit} + \psi_{ntji}|Q_i, Q_j) = \binom{n}{2}^{-1} \sum_{i=1}^n \sum_{j=1}^n E(\psi_{ntji}|Q_i, Q_j) \\
&= \binom{n}{2}^{-1} \sum_{i=1}^n \sum_{j=1}^n \frac{\epsilon_i \epsilon_j}{h^{2d+2}} \int SK\left(\frac{W_j - W_t}{h}\right) SK\left(\frac{W_i - W_t}{h}\right) \frac{1}{f(W_t)} dW_t.
\end{aligned}$$

$V(H_n^{(3)}(Q_{v_1}, Q_{v_2}, Q_{v_3})) = O_p(n^{-3} \sigma_{3n}^2)$ , where by A2, A3(1),

$$\sigma_{3n}^2 = E\phi_{nijt}^2 \leq CE\psi_{nijt}^2 = C \frac{1}{h^{4d+4}} E\left[\frac{1}{f^4(W_t)} SK^2\left(\frac{W_j - W_t}{h}\right) SK^2\left(\frac{W_i - W_t}{h}\right) \sigma^2(W_j) \sigma^2(W_t)\right] = O_p(h^{-2d-4}).$$

So  $V(H_n^{(3)}(Q_{v_1}, Q_{v_2}, Q_{v_3})) = O_p(n^{-3} h^{-2d-4})$ , and

$$H_n^{(3)}(Q_{v_1}, Q_{v_2}, Q_{v_3}) = O_p(n^{-\frac{3}{2}} h^{-d-2}) = O_p(n^{-1} h^{-\frac{d}{2}-2} n^{-\frac{1}{2}} h^{-\frac{d}{2}}) = o_p((nh^{2+\frac{d}{2}})^{-1}).$$

So in all, we obtain for  $i \neq j \neq t$ ,

$$\begin{aligned}
T_{211} \equiv T_{211e} &= \binom{n}{2}^{-1} \sum_{i=1}^n \sum_{j=1}^n \frac{\epsilon_i \epsilon_j}{h^{2d+2}} \int SK\left(\frac{W_j - W_t}{h}\right) SK\left(\frac{W_i - W_t}{h}\right) \frac{1}{f(W_t)} dW_t (1 + o_p(1)). \\
\text{(ii) } T_{211e} &= \binom{n}{2}^{-1} h^{-2-\frac{d}{2}} \sum_{i=1}^n \sum_{j=1}^n \frac{\epsilon_i \epsilon_j}{h^{\frac{3d}{2}}} \underbrace{\int SK\left(\frac{W_j - W_t}{h}\right) SK\left(\frac{W_i - W_t}{h}\right) \frac{1}{f(W_t)} dW_t}_{\phi_{nij}} (1 + o_p(1)).
\end{aligned}$$

Note that  $\phi_{nij} = \phi_n(Q_i, Q_j)$  is symmetric in  $Q_i$  and  $Q_j$ . Since  $E(\phi_{nij}|Q_i) = 0$ , by Theorem 1 in Hall (1984), if  $E\phi_{nij}^2 < \infty$ , and  $\frac{EG_n^2(Q_i, Q_j) + \frac{1}{n} E\phi_{nij}^4}{[E\phi_{nij}^2]^2} \rightarrow 0$ , then

$$nh^{2+\frac{d}{2}}T_{211e} = n \binom{n}{2}^{-1} \sum_{i=1}^n \sum_{\substack{j=1 \\ i < j}}^n \phi_{nij} \xrightarrow{d} N(0, 2E\phi_{nij}^2).$$

Recall the definition of  $SK(\cdot)$  in equation (1.5), and let  $\frac{W_i - W_t}{h} = \Psi$ ,  $\frac{W_j - W_i}{h} = \Psi_1$ , then  $\frac{W_j - W_t}{h} = \Psi_1 + \Psi$ ,  $W_t = W_i - h\Psi$ , and  $W_j = W_i + h\Psi_1$ . Then

$$\begin{aligned} 2E\phi_{nij}^2 &= 2E[\sigma^2(W_i)\sigma^2(W_j)\frac{1}{h^d}(\frac{1}{h^d} \int SK(\frac{W_j - W_t}{h})SK(\frac{W_i - W_t}{h})\frac{1}{f(W_t)}dW_t)^2] \\ &= 2 \int f(W_i)f(W_i + h\Psi_1)\sigma^2(W_i)\sigma^2(W_i + h\Psi_1)[\int \sum_{0 \leq |j| \leq p} \sum_{l=1}^{N_{|j|}} S^{-1} \sum_{l'=0}^{|j|-1} N_{l'} + l K(\Psi_1 + \Psi)(\Psi_1 + \Psi)^{G_{|j|}(l)} \\ &\quad \sum_{0 \leq |j'| \leq p} \sum_{l'=1}^{N_{|j'|}} S^{-1} \sum_{l''=0}^{|j'|-1} N_{l''} + l' K(\Psi)\Psi^{G_{|j'|}(l')} \frac{1}{f(W_i - h\Psi)}d\Psi]^2 dW_i d\Psi_1 \\ &\rightarrow 2[\sigma^4(W_i)dW_i][\int (\int SK(\Psi_1 + \Psi)SK(\Psi)d\Psi)^2 d\Psi_1] = \Omega. \end{aligned}$$

$$E\phi_{nij}^4 = h^{-d} \int E(\epsilon_i^4|W_i)E(\epsilon_j^4|W_i + h\Psi_1)[\int SK(\Psi_1 + \Psi)SK(\Psi)\frac{1}{f(W_i - h\Psi)}d\Psi]^4 f(W_i)f(W_i + h\Psi_1)dW_i d\Psi_1,$$

thus, with A1, A2, and A3(1),  $\frac{1}{n}E\phi_{nij}^4 \rightarrow 0$  as  $nh^d \rightarrow \infty$ .

With similar change of variables, i.e.,  $\Psi = \frac{W_i - W_t}{h}$ ,  $\Psi_1 = \frac{W_j - W_t}{h}$ ,  $\Psi_2 = \frac{W_i - W_j}{h}$ ,  $\Psi_3 = \frac{W_i - W_j}{h}$ , we have

$$\begin{aligned} EG_n^2(Q_i, Q_j) &= E[E^2(\phi_{ni}\phi_{nj}|Q_i, Q_j)] \\ &= h^d E\{\epsilon_i^2 \epsilon_j^2 [\int \sigma^2(W_i)(\frac{1}{h^d} \int SK(\frac{W_i - W_t}{h})SK(\frac{W_i - W_t}{h})\frac{1}{f(W_t)}dW_t) \\ &\quad (\frac{1}{h^d} \int SK(\frac{W_j - W_t}{h})SK(\frac{W_t - W_t}{h})\frac{1}{f(W_t)}dW_t')f(W_i)dW_i]^2\} \\ &= h^{4d} \int \sigma^2(W_i)\sigma^2(W_i - h\Psi_3)f(W_i)f(W_i - h\Psi_3)[\int \sigma^2(W_i + h\Psi_2)(\int SK(\Psi)SK(\Psi_2 + \Psi)\frac{1}{f(W_i - h\Psi)}d\Psi) \\ &\quad (\int SK(\Psi_1)SK(\Psi_2 + \Psi_1 + \Psi_3)\frac{1}{f(W_i - h\Psi_3 - h\Psi_1)}d\Psi_1)f(W_i + h\Psi_2)d\Psi_2]^2 d\Psi_3 dW_i \\ &= O(h^{4d}), \end{aligned}$$

so we verify the condition for the asymptotic normality and we have  $nh^{2+\frac{d}{2}}T_{211e} \xrightarrow{d} N(0, \Omega)$ .

(iii) Since  $i \neq j$  and  $i \neq t$ , the only other case is for  $i \neq j = t$ . In this case, since the kernel is symmetric,

$$T_{211} \equiv T_{211B} = \frac{1}{n^3 h^{2d+2}} \sum_{i=1}^n \sum_{\substack{j=1 \\ i \neq j}}^n \frac{\epsilon_j^2}{f^2(W_i)} SK^2(\frac{W_j - W_i}{h}) = \frac{1}{2nh^{2+d}} \frac{2}{n^2} \sum_{i=1}^n \sum_{\substack{j=1 \\ i \neq j}}^n \underbrace{\frac{\epsilon_j^2}{h^d f^2(W_i)} SK^2(\frac{W_j - W_i}{h})}_{\psi_{nij}}.$$

Note that  $\frac{2}{n^2} \sum_{i=1}^n \sum_{\substack{j=1 \\ i \neq j}}^n \psi_{nij} = [O(n^{-3}) + \binom{n}{2}^{-1} \sum_{i=1}^n \sum_{\substack{j=1 \\ i \neq j}}^n \phi_{nij}]$  for  $\phi_{nij} = \psi_{nij} + \psi_{nji}$ . We apply Lemma

2 here again and write  $\binom{n}{2}^{-1} \sum_{i=1}^n \sum_{\substack{j=1 \\ i \neq j}}^n \phi_{nij} = \theta_n + \sum_{j=1}^2 \binom{2}{j} H_n^{(j)}(Q_{v_1}, \dots, Q_{v_j})$ .

$$\theta_n = E\phi_{nij} = 2 \int \frac{1}{h^d f^2(W_i)} SK^2(\frac{W_j - W_i}{h}) \sigma^2(W_j) f(W_i) f(W_j) dW_i dW_j \rightarrow 2 \int \sigma^2(W_i) dW_i \int SK^2(\Psi) d\Psi.$$

$$H_n^{(2)}(Q_{v_1}, Q_{v_2}) = O_p(n^{-1} \sigma_{2n}) = O_p(n^{-1} h^{-\frac{d}{2}}), \text{ since } \sigma_{2n}^2 = E\phi_{nij}^2 = O(h^{-d}).$$

$$H_n^{(1)}(Q_{v_1}) = O_p(n^{-\frac{1}{2}} \sigma_{1n}) = O_p(n^{-\frac{1}{2}}), \text{ since } \sigma_{1n}^2 = E[E^2(\phi_{nij}|Q_i)] = E[\int \frac{1}{f^2(W_i)} SK^2(\Psi) \sigma^2(W_i + h\Psi) d\Psi]^2 = O(1).$$

So in all,  $T_{211B} = \frac{1}{nh^{2+d}} \int \sigma^2(W_i) dW_i \int SK^2(\Psi) d\Psi (1 + o_p(1)) = \frac{1}{nh^{2+d}} B_1 (1 + o_p(1))$ .

Combing above (i)-(iii), we obtain that  $nh^{2+\frac{d}{2}}(T_{211} - \frac{1}{nh^{2+d}} B_1) \xrightarrow{d} N(0, \Omega)$ .

(b)

$$\begin{aligned} T_{212} &= \frac{1}{n^3} \sum_{i=1}^n \sum_{j=1}^n \sum_{t=1}^n \frac{1}{f^2(W_i)h^{2d+2}} SK\left(\frac{W_j - W_i}{h}\right) SK\left(\frac{W_t - W_i}{h}\right) \epsilon_j \sum_{|\mathbf{k}|=p+1} \frac{h^{p+1}}{\mathbf{k}!} (D^{\mathbf{k}}m)(W_i + \lambda(W_t - W_i)) \left(\frac{W_t - W_i}{h}\right)^{\mathbf{k}} \\ &= \frac{1}{n^3} \sum_{i=1}^n \sum_{j=1}^n \sum_{t=1}^n \psi_{nijt}. \end{aligned}$$

(i) Consider the case  $i \neq j \neq t$ . With  $\phi_{nijt} = \psi_{nijt}\psi_{nitj} + \psi_{njit} + \psi_{njti} + \psi_{ntij} + \psi_{ntji}$ , we apply Lemma

$$2 \text{ to obtain } T_{212} = \frac{1}{6} \binom{n}{3}^{-1} \sum_{i=1}^n \sum_{j=1}^n \sum_{t=1}^n \phi_{nijt} (1 + o_p(1)) = \frac{1}{6} [\theta_n + \sum_{j=1}^3 \binom{3}{j} H_n^{(j)}(Q_{v_1}, \dots, Q_{v_j})].$$

Here  $\theta_n = 0$  as  $E(\epsilon_j | W_j) = 0$ .

$$\begin{aligned} H_n^{(1)}(Q_{v_1}) &= n^{-1} \sum_{i=1}^n E(\phi_{nijt} | Q_i) \\ &= \frac{1}{n} \sum_{i=1}^n \frac{2\epsilon_i}{h^{2d+2}} \int \frac{1}{f^2(W_j)} SK\left(\frac{W_i - W_j}{h}\right) SK\left(\frac{W_t - W_j}{h}\right) \sum_{|\mathbf{k}|=p+1} \frac{h^{p+1}}{\mathbf{k}!} (D^{\mathbf{k}}m)(W_j + \lambda(W_t - W_j)) \left(\frac{W_t - W_j}{h}\right)^{\mathbf{k}} \\ &\quad \times f(W_j) f(W_t) dW_j dW_t \\ &= O_p(n^{-\frac{1}{2}} h^{p-1}), \end{aligned}$$

since  $(D^{\mathbf{k}}m)(W)$  is bounded and uniformly continuous for  $|\mathbf{k}| = p+1$  as in A3(2).

Given that  $\sigma_{2n}^2 = E(\phi_{2n}^2(Q_i, Q_j))$ ,  $\phi_{2n}(Q_i, Q_j) = E(\phi_{nijt} | Q_i, Q_j) = E(\psi_{nijt} + \psi_{njit} + \psi_{ntij} + \psi_{ntji} | Q_i, Q_j)$ ,

$$E(\psi_{nijt} | Q_i, Q_j) = \frac{\epsilon_j}{f^2(W_i)h^{2d+2}} \int SK\left(\frac{W_j - W_i}{h}\right) SK\left(\frac{W_t - W_i}{h}\right) \sum_{|\mathbf{k}|=p+1} \frac{h^{p+1}}{\mathbf{k}!} (D^{\mathbf{k}}m)(W_i + \lambda(W_t - W_i)) \left(\frac{W_t - W_i}{h}\right)^{\mathbf{k}} f(W_t) dW_t.$$

We obtain  $E[E^2(\psi_{nijt} | Q_i, Q_j)] = O(h^{2p+2-4-d}) = O(h^{2p-d-2})$ . Since similar arguments apply to the other terms, we have  $\sigma_{2n}^2 = O(h^{2p-d-2})$ . Since  $V(H_n^{(2)}(Q_{v_1}, Q_{v_2})) = O(n^{-2}\sigma_{2n}^2)$ , we conclude

$$H_n^{(2)}(Q_{v_1}, Q_{v_2}) = O_p(n^{-1} h^{p-\frac{d}{2}-1}) = o_p(n^{-1} h^{-\frac{d}{2}-2}).$$

$$\sigma_{3n}^2 = E(\phi_{nijt}^2) = O(h^{2(p-d)-2}), V(H_n^{(3)}(Q_{v_1}, Q_{v_2}, Q_{v_3})) = O_p(n^{-3}\sigma_{3n}^2), \text{ and thus since } nh^d \rightarrow \infty,$$

$$H_n^{(3)}(Q_{v_1}, Q_{v_2}, Q_{v_3}) = O_p(n^{-\frac{3}{2}} p h^{(p-d)-1}) = O_p((n^{-1} h^{-\frac{d}{2}-2})(n^{-\frac{1}{2}} h^{(p-\frac{d}{2})+1}) = o_p((n^{-1} h^{-\frac{d}{2}-2}).$$

So  $T_{212} = O_p(n^{-\frac{1}{2}} h^{p-1}) + o_p((n^{-1} h^{-\frac{d}{2}-2}) = o_p((n^{-1} h^{-\frac{d}{2}-2})$  with  $nh^{2p+d+2} \rightarrow 0$  as in A4(2).

(ii) When  $i \neq j = t$ , then

$$T_{212} = \frac{1}{n^3} \sum_{i=1}^n \sum_{j=1}^n \frac{1}{f^2(W_i)h^{2d+2}} SK^2\left(\frac{W_j - W_i}{h}\right) \epsilon_j \sum_{|\mathbf{k}|=p+1} \frac{h^{p+1}}{\mathbf{k}!} (D^{\mathbf{k}}m)(W_i + \lambda(W_j - W_i)) \left(\frac{W_j - W_i}{h}\right)^{\mathbf{k}}.$$

We again apply Lemma to obtain with similar arguments that

$$T_{212} = O_p(n^{-2} h^{p-\frac{3}{2}d-1}) + O_p(n^{-\frac{3}{2}} h^{p-d-1}) = O_p(n^{-\frac{3}{2}} h^{p-d-1}) = O_p(n^{-1} h^{-\frac{d}{2}-2} n^{-\frac{1}{2}} h^{p-\frac{d}{2}+1}) = o_p(n^{-1} h^{-\frac{d}{2}-2}).$$

(i)-(ii) above give the results that  $T_{212} = o_p(n^{-1} h^{-\frac{d}{2}-2})$ .

(c)  $T_{213} = o_p(n^{-1} h^{-\frac{d}{2}-2})$  follows from the same arguments in (b) for  $T_{212}$ .

(d)

$$\begin{aligned}
T_{214} &= \frac{1}{n^3} \sum_{i=1}^n \sum_{j=1}^n \sum_{t=1}^n \frac{1}{f^2(W_i)h^{2d+2}} SK\left(\frac{W_j - W_i}{h}\right) SK\left(\frac{W_t - W_i}{h}\right) \sum_{|\mathbf{k}|=p+1} \frac{h^{p+1}}{\mathbf{k}!} (D^{\mathbf{k}}m)(W_i + \lambda(W_j - W_i)) \left(\frac{W_j - W_i}{h}\right)^{\mathbf{k}} \\
&\quad \times \sum_{|\mathbf{k}'|=p+1} \frac{h^{p+1}}{\mathbf{k}'!} (D^{\mathbf{k}'}m)(W_i + \lambda(W_t - W_i)) \left(\frac{W_t - W_i}{h}\right)^{\mathbf{k}'} \\
&= \frac{1}{n^3} \sum_{i=1}^n \sum_{j=1}^n \sum_{t=1}^n \psi_{nijt}.
\end{aligned}$$

(i)  $i \neq j \neq t$ . Note that  $\psi_{nijt}$  is symmetric in  $j, t$ . So we let  $\phi_{nijt} = \psi_{nijt} + \psi_{njit} + \psi_{ntji}$  and apply

Lemma 2 to have

$$T_{214} = \frac{1}{3} \binom{n}{3}^{-1} \sum_{i=1}^n \sum_{j=1}^n \sum_{t=1}^n \phi_{nijt} (1 + o_p(1)) = (1 + o_p(1))^{\frac{1}{3}} [\theta_n + \sum_{j=1}^3 \binom{3}{j} H_n^{(j)}(Q_{v_1}, \dots, Q_{v_j})].$$

Recall the definition of  $SK(\cdot)$  in equation (1.5),

$$\begin{aligned}
\frac{1}{3}\theta_n &= E\psi_{nijt} \\
&= h^{2p} \sum_{0 \leq |\mathbf{j}| \leq p} \sum_{l=1}^N \sum_{|\mathbf{j}'| \leq p} \sum_{l'=1}^N S^{-1} \int \frac{1}{f^2(W_i)} K(\Psi_1) \\
&\quad \times \sum_{|\mathbf{k}|=p+1} \frac{h^{p+1}}{\mathbf{k}!} (D^{\mathbf{k}}m)(W_i + \lambda h \Psi_1)(\Psi_1)^{\mathbf{k}+G_{|\mathbf{j}|}(l)} K(\Psi_2) \sum_{|\mathbf{k}'|=p+1} \frac{h^{p+1}}{\mathbf{k}'!} (D^{\mathbf{k}'}m)(W_i + \lambda h \Psi_2)(\Psi_2)^{\mathbf{k}'+G_{|\mathbf{j}'|}(l')} \\
&\quad \times f(W_i) f(W_i + h \Psi_1) f(W_i + h \Psi_2) dW_i d\Psi_1 d\Psi_2 \\
&= O(h^{2p}) = o(n^{-1}h^{-\frac{d}{2}-2}) \text{ since } nh^{\frac{d}{2}+2}h^{2p} \rightarrow 0 \text{ in A4(2).}
\end{aligned}$$

$$H_n^{(1)}(Q_{v_1}) = O_p(n^{-\frac{1}{2}}h^{2p}) = o_p(n^{-1}h^{-\frac{d}{2}-2}) \text{ since } V(H_n^{(1)}(Q_{v_1})) = O(n^{-1}\sigma_{1n}^2) \text{ and } \sigma_{1n}^2 = O(h^{4p}).$$

$$H_n^{(2)}(Q_{v_1}, Q_{v_2}) = O_p(n^{-1}h^{2p-\frac{d}{2}}) = o_p(n^{-1}h^{-\frac{d}{2}-2}) \text{ since } V(H_n^{(2)}(Q_{v_1}, Q_{v_2})) = O(n^{-2}\sigma_{2n}^2) \text{ and } \sigma_{2n}^2 =$$

$O(h^{4p-d})$ .

$$H_n^{(3)}(Q_{v_1}, Q_{v_2}, Q_{v_3}) = O_p(n^{-\frac{3}{2}}h^{2p-d}) = o_p(n^{-1}h^{-\frac{d}{2}-2}) \text{ since } V(H_n^{(3)}(Q_{v_1}, Q_{v_2}, Q_{v_3})) = O(n^{-3}\sigma_{3n}^2)$$

and  $\sigma_{3n}^2 = O(h^{4p-2d})$ .

So when  $i \neq j \neq t$ ,  $T_{214} = o_p(n^{-1}h^{-\frac{d}{2}-2})$ .

(ii) When  $i \neq j = t$ ,

$$T_{214} = \frac{1}{n^3} \sum_{i=1}^n \sum_{j=1}^n \sum_{t=1}^n \frac{1}{f^2(W_i)h^{2d+2}} SK^2\left(\frac{W_j - W_i}{h}\right) \left( \sum_{|\mathbf{k}|=p+1} \frac{h^{p+1}}{\mathbf{k}!} (D^{\mathbf{k}}m)(W_i + \lambda(W_j - W_i)) \left(\frac{W_j - W_i}{h}\right)^{\mathbf{k}} \right)^2.$$

$$\text{Since } E|T_{214}| = O(n^{-1}h^{2(p+1)-d-2}) = O(n^{-1}h^{2p-d}) = o_p(n^{-1}h^{-\frac{d}{2}-2}h^{2p-\frac{d}{2}+2}) = o_p(n^{-1}h^{-\frac{d}{2}-2}),$$

where we need  $p+1 > \frac{d}{4}$  in A4(2), we obtain that  $T_{214} = o_p(n^{-1}h^{-\frac{d}{2}-2})$ .

In all,  $T_{214} = o_p(n^{-1}h^{-\frac{d}{2}-2})$ .

Combining results in (1)(a)-(d), we obtain the claim in (1).

(2)  $T_{23}$  calls for

$$\begin{aligned}
&\hat{g}_x(\omega_l; \omega_i^c) - g_x(\omega_l; \omega_i^c) \\
&= \frac{1}{f(\omega_l; \omega_i^c)nh^{d+1}} \sum_{j=1}^n SK\left(\frac{\omega_j - \omega_l}{h}; \frac{\omega_j^c - \omega_i^c}{h}\right) \left[ \sum_{|\mathbf{k}|=p+1} \frac{h^{p+1}}{\mathbf{k}!} (D^{\mathbf{k}}m)(\omega_l + \lambda(\omega_j - \omega_l); \omega_i^c + \lambda(\omega_j^c - \omega_i^c)) \left(\frac{\omega_j - \omega_l}{h}; \frac{\omega_j^c - \omega_i^c}{h}\right)^{\mathbf{k}} \right. \\
&\quad \left. + \epsilon_j \right] (1 + o_p(1)),
\end{aligned}$$

where we follow equation (1.5) to define

$$SK\left(\frac{\omega_j - \omega_l}{h}; \frac{\omega_j^c - \omega_i^c}{h}\right) = \sum_{0 \leq |\mathbf{k}| \leq pm=1} \sum_{1+d, \sum_{m'=0}^{|\mathbf{k}|-1} N_{m'+m}} S^{-1} K\left(\frac{\omega_j - \omega_l}{h}; \frac{\omega_j^c - \omega_i^c}{h}\right) \left(\frac{\omega_j - \omega_l}{h}; \frac{\omega_j^c - \omega_i^c}{h}\right)^{G_{|\mathbf{k}|}(m)}. \quad (1.13)$$

Here we write  $(\frac{W_j - W_i}{h}) \equiv (\frac{\omega_j - \omega_l}{h}; \frac{\omega_j^c - \omega_i^c}{h})$  to denote an equivalent rearrangement, highlighting the variables  $\omega$  in  $W$  to be tested for significance.

With equations (1.5) and (1.13), we have

$$\begin{aligned} T_{23} &= -\frac{2}{n} \sum_{i=1}^n \left[ \frac{1}{f(W_i)nh^{d+1}} \sum_{j=1}^n SK\left(\frac{W_j - W_i}{h}\right) \left( \sum_{|\mathbf{k}|=p+1} \frac{h^{p+1}}{\mathbf{k}!} (D^{\mathbf{k}}m)(W_i + \lambda(W_j - W_i)) \left(\frac{W_j - W_i}{h}\right)^{\mathbf{k}} + \epsilon_j \right) \right. \\ &\quad \times \frac{1}{n} \sum_{l=1}^n \frac{1}{f(\omega_l; \omega_i^c)nh^{d+1}} \sum_{t=1}^n SK\left(\frac{\omega_t - \omega_l}{h}; \frac{\omega_i^c - \omega_i^c}{h}\right) \left[ \sum_{|\mathbf{k}|=p+1} \frac{h^{p+1}}{\mathbf{k}!} (D^{\mathbf{k}}m)(\omega_l + \lambda(\omega_t - \omega_l); \omega_i^c + \lambda(\omega_t^c - \omega_i^c)) \right. \\ &\quad \times \left. \left(\frac{\omega_t - \omega_l}{h}; \frac{\omega_i^c - \omega_i^c}{h}\right)^{\mathbf{k}} + \epsilon_t \right] \\ &= -\frac{2}{n^4} \sum_{i=1}^n \sum_{j=1}^n \sum_{l=1}^n \sum_{t=1}^n \frac{1}{h^{2d+2}f(W_i)f(\omega_l; \omega_i^c)} SK\left(\frac{W_j - W_i}{h}\right) SK\left(\frac{\omega_t - \omega_l}{h}; \frac{\omega_i^c - \omega_i^c}{h}\right) [\epsilon_j \epsilon_t \\ &\quad + \epsilon_t \sum_{|\mathbf{k}|=p+1} \frac{h^{p+1}}{\mathbf{k}!} (D^{\mathbf{k}}m)(W_i + \lambda(W_j - W_i)) \left(\frac{W_j - W_i}{h}\right)^{\mathbf{k}} \\ &\quad + \epsilon_j \sum_{|\mathbf{k}|=p+1} \frac{h^{p+1}}{\mathbf{k}!} (D^{\mathbf{k}}m)(\omega_l + \lambda(\omega_t - \omega_l); \omega_i^c + \lambda(\omega_t^c - \omega_i^c)) \left(\frac{\omega_t - \omega_l}{h}; \frac{\omega_i^c - \omega_i^c}{h}\right)^{\mathbf{k}} \\ &\quad + \sum_{|\mathbf{k}|=p+1} \frac{h^{p+1}}{\mathbf{k}!} (D^{\mathbf{k}}m)(W_i + \lambda(W_j - W_i)) \left(\frac{W_j - W_i}{h}\right)^{\mathbf{k}} \\ &\quad \times \sum_{|\mathbf{k}|=p+1} \frac{h^{p+1}}{\mathbf{k}!} (D^{\mathbf{k}}m)(\omega_l + \lambda(\omega_t - \omega_l); \omega_i^c + \lambda(\omega_t^c - \omega_i^c)) \left(\frac{\omega_t - \omega_l}{h}; \frac{\omega_i^c - \omega_i^c}{h}\right)^{\mathbf{k}}] \\ &= T_{231} + T_{232} + T_{233} + T_{234}. \end{aligned}$$

(a) We claim that

$$\begin{aligned} T_{231} &= -\frac{2}{n^4} \sum_{i=1}^n \sum_{j=1}^n \sum_{l=1}^n \sum_{t=1}^n \frac{1}{h^{2d+2}f(W_i)f(\omega_l; \omega_i^c)} SK\left(\frac{W_j - W_i}{h}\right) SK\left(\frac{\omega_t - \omega_l}{h}; \frac{\omega_i^c - \omega_i^c}{h}\right) \epsilon_j \epsilon_t = -\frac{2}{n^4} \sum_{i=1}^n \sum_{j=1}^n \sum_{l=1}^n \sum_{t=1}^n \psi_{nijlt} \\ &= -\frac{2}{nh^{2+d-1}} B_3(1 + o_p(1)). \text{ The claim follows from results (i)-(iii) below.} \end{aligned}$$

(i) When  $i \neq j \neq l \neq t$ , we apply Lemma 2 to have

$$T_{231} = -\frac{2}{24} \binom{n}{4}^{-1} \sum_{i=1}^n \sum_{j=1}^n \sum_{l=1}^n \sum_{t=1}^n \psi_{nijlt} (1 + o_p(1)) = -\frac{2}{24} \binom{n}{4}^{-1} \sum_{i=1}^n \sum_{j=1}^n \sum_{l=1}^n \sum_{t=1}^n \phi_{nijlt} (1 + o_p(1)),$$

where

$$\begin{aligned} \phi_{nijlt} &= \psi_{nijlt} + \psi_{nijtl} + \psi_{niltj} + \psi_{niljt} + \psi_{nitjl} + \psi_{nitlj} + \psi_{njilt} + \psi_{njitl} + \psi_{njlti} + \psi_{njlit} + \psi_{njtli} + \psi_{njtil} \\ &\quad + \psi_{nlijt} + \psi_{nlitj} + \psi_{nljti} + \psi_{nljit} + \psi_{nltji} + \psi_{nltij} + \psi_{ntijl} + \psi_{ntilj} + \psi_{ntjli} + \psi_{ntjl i} + \psi_{ntlji} + \psi_{ntlji} \end{aligned}$$

present a permutation of  $\psi_{nijlt}$  across the indices, so that  $\phi_{nijlt}$  is symmetric in the indices. Again, we

write

$$T_{231} = -\frac{2}{24} [\theta_n + \sum_{j=1}^4 \binom{4}{j}] H_n^{(j)}(Q_{v_1}, \dots, Q_{v_j}).$$

$V(H_n^{(2)}(Q_{v_1}, Q_{v_2})) = O(n^{-2}\sigma_{2n}^2)$ .  $\sigma_{2n}^2 = V(\phi_{2n}(Q_1, Q_2))$ ,  $\phi_{2n}(Q_j, Q_t) = E(\phi_{nijlt}|Q_j, Q_t) = E(\psi_{nijlt} + \psi_{nljit} + \psi_{nitlj} + \psi_{nltij}|Q_j, Q_t)$ . Since



$E(\psi_{nijlt}|Q_j, Q_t) = \epsilon_t \epsilon_j \int \frac{1}{h^{2d+2} f(W_i) f(\omega_l; \omega_i^c)} SK(\frac{W_j - W_i}{h}) SK(\frac{\omega_t - \omega_l}{h}; \frac{\omega_i^c - \omega_i}{h}) f(W_i) f(W_l) dW_i dW_l$ ,  
 $E[E^2(\psi_{nijlt}|Q_j, Q_t)] = O(h^{-4-(d-d_1)})$ . Similar arguments apply to the other terms and we obtain  $\sigma_{2n}^2 = O(h^{-4-(d-d_1)})$ . So  $H_n^{(2)}(Q_{v_1}, Q_{v_2}) = O_p(n^{-1}h^{-2-\frac{(d-d_1)}{2}}) = o_p(n^{-1}h^{-2-\frac{d}{2}})$ , since  $d_1 \geq 1$ .  
 $V(H_n^{(3)}(Q_{v_1}, Q_{v_2}, Q_{v_3})) = O(n^{-3}\sigma_{3n}^2)$ .  $\sigma_{3n}^2 = V(\phi_{3n}(Q_1, Q_2, Q_3))$ ,  $\phi_{3n}(Q_i, Q_j, Q_l) = E(\phi_{nijlt}|Q_i, Q_j, Q_l) = E(\psi_{ntjli} + \psi_{ntjil} + \psi_{ntlij} + \psi_{ntlji} + \psi_{ntilj} + \psi_{ntilj} + \psi_{nijtl} + \psi_{njitl} + \psi_{njlti} + \psi_{nljti} + \psi_{niltj} + \psi_{nilitj}|Q_j, Q_t)$ .

Since

$E(\psi_{ntjli}|Q_i, Q_j, Q_l) = \epsilon_i \epsilon_j \int \frac{1}{h^{2d+2} f(W_t) f(\omega_l; \omega_i^c)} SK(\frac{W_j - W_t}{h}) SK(\frac{\omega_i - \omega_l}{h}; \frac{\omega_i^c - \omega_i}{h}) f(W_t) dW_t$ ,  
 $E[E^2(\psi_{ntjli}|Q_i, Q_j, Q_l)] = O(h^{-4-d})$ . Similar arguments apply to the other terms and we obtain  $\sigma_{3n}^2 = O(h^{-4-d})$ . So  $H_n^{(3)}(Q_{v_1}, Q_{v_2}, Q_{v_3}) = O_p(n^{-\frac{3}{2}}h^{-2-\frac{d}{2}}) = o_p(n^{-1}h^{-2-\frac{d}{2}})$ .

$V(H_n^{(4)}(Q_{v_1}, Q_{v_2}, Q_{v_3}, Q_{v_4})) = O(n^{-4}\sigma_{4n}^2)$ ,  $\sigma_{4n}^2 \leq CE\psi_{nijlt}^2 = O_p(h^{-4-2d})$  by A2, A3(1), and A1(2),

so we obtain

$$H_n^{(4)}(Q_{v_1}, Q_{v_2}, Q_{v_3}, Q_{v_4}) = O_p(n^{-2}h^{-2-d}) = o_p(n^{-1}h^{-2-\frac{d}{2}}).$$

So in all,  $T_{231} = o_p(n^{-1}h^{-2-\frac{d}{2}})$  when  $i \neq j \neq l \neq t$ .

(ii) When  $i \neq j = l \neq t$ ,

$$T_{231} = -\frac{2}{n^4} \sum_{i=1}^n \sum_{j=1}^n \sum_{t=1}^n \frac{1}{h^{2d+2} f(W_i) f(\omega_j; \omega_i^c)} SK(\frac{W_j - W_i}{h}) SK(\frac{\omega_t - \omega_l}{h}; \frac{\omega_i^c - \omega_i}{h}) \epsilon_t \epsilon_j = O_p(n^{-1}h^{-2}) = o_p(n^{-1}h^{-2-\frac{d}{2}}).$$

(iii) When  $i \neq j = t \neq l$ ,

$$\begin{aligned} T_{231} &\equiv T_{231B} = -\frac{2}{(nh^{2+d-d_1})6} \frac{6}{n^3} \sum_{i=1}^n \sum_{j=1}^n \sum_{l=1}^n \underbrace{\frac{1}{h^{2d+2} f(W_i) f(\omega_l; \omega_i^c)} SK(\frac{W_j - W_i}{h}) SK(\frac{\omega_j - \omega_l}{h}; \frac{\omega_j^c - \omega_i}{h}) \epsilon_j^2}_{\psi_{nijl}} \\ &= -\frac{2}{nh^{2+d-d_1}} B_3(1 + o_p(1)). \\ \text{As } \frac{6}{n^3} \sum_{i=1}^n \sum_{j=1}^n \sum_{l=1}^n \psi_{nijl} &= \binom{n}{3}^{-1} \sum_{i=1}^n \sum_{j=1}^n \sum_{l=1}^n \phi_{nijl} = \theta_n + \sum_{j=1}^3 \binom{3}{j} H_n^{(j)}(Q_{v_1}, \dots, Q_{v_j}) \text{ for } \phi_{nijl} = \psi_{nijl} + \\ &\psi_{nilj} + \psi_{njil} + \psi_{njli} + \psi_{nlij} + \psi_{nli j}. \text{ Since} \end{aligned}$$

$$\theta_n = 6E\psi_{nijl} \rightarrow 6 \int \sigma^2(W_i) f(\omega_i; \omega_i^c) dW_i d\omega_i^c \int SK(\Psi_1) SK(\psi_{21}; \psi_{12}) d\psi_1 d\psi_{21} = 6B_3,$$

$$H_n^{(1)}(Q_{v_1}) = O_p(n^{-\frac{1}{2}}\sigma_{1n}) = O_p(n^{-\frac{1}{2}}) = o_p(1),$$

$$H_n^{(2)}(Q_{v_1}, Q_{v_2}) = O_p(n^{-1}\sigma_{2n}) = O_p(n^{-1}h^{-\frac{d}{2}}) = o_p(1),$$

$$H_n^{(3)}(Q_{v_1}, Q_{v_2}, Q_{v_3}) = O_p(n^{-\frac{3}{2}}\sigma_{3n}) = O_p(n^{-\frac{3}{2}}h^{-\frac{d+d_1}{2}}) = o_p(1),$$

we have  $\frac{6}{n^3} \sum_{i=1}^n \sum_{j=1}^n \sum_{l=1}^n \psi_{nijl} = 6B_3 + o_p(1)$  and the claim above.

With similar but lengthy arguments, we show that

$$\begin{aligned} &(b) T_{232} \\ &= -\frac{2}{n^4} \sum_{i=1}^n \sum_{j=1}^n \sum_{l=1}^n \sum_{t=1}^n \frac{1}{h^{2d+2} f(W_i) f(\omega_l; \omega_i^c)} SK(\frac{W_j - W_i}{h}) SK(\frac{\omega_t - \omega_l}{h}; \frac{\omega_i^c - \omega_i}{h}) \epsilon_t \sum_{|\mathbf{k}|=p+1} \frac{h^{p+1}}{\mathbf{k}!} (D^{\mathbf{k}} m)(W_i + \lambda(W_j - W_i)) \\ &\quad \times (\frac{W_j - W_i}{h})^{\mathbf{k}} \\ &= o_p(n^{-1}h^{-2-\frac{d}{2}}). \end{aligned}$$

$$\begin{aligned}
& (c) T_{233} \\
&= -\frac{2}{n^4} \sum_{i=1}^n \sum_{j=1}^n \sum_{l=1}^n \sum_{t=1}^n \frac{1}{h^{2d+2} f(W_i) f(\omega_l; \omega_i^c)} SK\left(\frac{W_j - W_i}{h} SK\left(\frac{\omega_t - \omega_l}{h}; \frac{\omega_t^c - \omega_i^c}{h}\right) \epsilon_j\right. \\
&\quad \times \sum_{\substack{\mathbf{k}! \\ |\mathbf{k}|=p+1}} \frac{h^{p+1}}{\mathbf{k}!} (D^{\mathbf{k}} m)(\omega_l + \lambda(\omega_t - \omega_l); \omega_i^c + \lambda(\omega_t^c - \omega_i^c)) \left(\frac{\omega_t - \omega_l}{h}; \frac{\omega_t^c - \omega_i^c}{h}\right)^{\mathbf{k}} \\
&= o_p(n^{-1} h^{-2-\frac{d}{2}}). \\
& (d) T_{234} \\
&= -\frac{2}{n^4} \sum_{i=1}^n \sum_{j=1}^n \sum_{l=1}^n \sum_{t=1}^n \frac{1}{h^{2d+2} f(W_i) f(\omega_l; \omega_i^c)} SK\left(\frac{W_j - W_i}{h} SK\left(\frac{\omega_t - \omega_l}{h}; \frac{\omega_t^c - \omega_i^c}{h}\right) \sum_{|\mathbf{k}|=p+1} \frac{h^{p+1}}{\mathbf{k}!} (D^{\mathbf{k}} m)(W_i + \lambda(W_j - W_i))\right. \\
&\quad \times \left(\frac{W_j - W_i}{h}\right)^{\mathbf{k}} \sum_{|\mathbf{k}|=p+1} \frac{h^{p+1}}{\mathbf{k}!} (D^{\mathbf{k}} m)(\omega_l + \lambda(\omega_t - \omega_l); \omega_i^c + \lambda(\omega_t^c - \omega_i^c)) \left(\frac{\omega_t - \omega_l}{h}; \frac{\omega_t^c - \omega_i^c}{h}\right)^{\mathbf{k}} \\
&= o_p(n^{-1} h^{-2-\frac{d}{2}}).
\end{aligned}$$

The claim in (2) follows from (2)(a)-(d) above.

$$(3) \text{ Following (2) above, we let } D^{\mathbf{k}} m_{tj;ti} = \sum_{|\mathbf{k}|=p+1} \frac{h^{p+1}}{\mathbf{k}!} (D^{\mathbf{k}} m)(\omega_j + \lambda(\omega_t - \omega_j); \omega_i^c + \lambda(\omega_t^c - \omega_i^c)) \left(\frac{\omega_t - \omega_j}{h}; \frac{\omega_t^c - \omega_i^c}{h}\right)^{\mathbf{k}},$$

and write

$$\begin{aligned}
T_{22} &= \frac{1}{n} \sum_{i=1}^n \left[ \frac{1}{n} \sum_{j=1}^n (\hat{g}_x(\omega_j; \omega_i^c) - g_x(\omega_j; \omega_i^c)) \right]^2 \\
&= \frac{1}{n} \sum_{i=1}^n \left[ \frac{1}{n} \sum_{j=1}^n \frac{1}{f(\omega_j; \omega_i^c) n h^{d+1}} \sum_{t=1}^n SK\left(\frac{\omega_t - \omega_j}{h}; \frac{\omega_t^c - \omega_i^c}{h}\right) [D^{\mathbf{k}} m_{tj;ti} + \epsilon_t]^2 \right. \\
&= \frac{1}{n^5} \sum_{i=1}^n \sum_{j=1}^n \sum_{t=1}^n \sum_{l=1}^n \sum_{m=1}^n \frac{1}{h^{2d+2} f(\omega_j; \omega_i^c) f(\omega_l; \omega_i^c)} SK\left(\frac{\omega_t - \omega_j}{h}; \frac{\omega_t^c - \omega_i^c}{h}\right) SK\left(\frac{\omega_m - \omega_l}{h}; \frac{\omega_m^c - \omega_i^c}{h}\right) [\epsilon_t \epsilon_m \\
&\quad + \epsilon_t \sum_{|\mathbf{k}|=p+1} h^{p+1} D^{\mathbf{k}} m_{ml;mi} + \epsilon_m \sum_{|\mathbf{k}|=p+1} h^{p+1} D^{\mathbf{k}} m_{tj;ti} \\
&\quad + h^{2(p+1)} \sum_{|\mathbf{k}|=p+1} h^{p+1} D^{\mathbf{k}} m_{tj;ti} \sum_{|\mathbf{k}|=p+1} h^{p+1} D^{\mathbf{k}} m_{ml;mi}] \\
&= T_{221} + T_{222} + T_{223} + T_{224}. \\
(a) T_{221} &= \frac{1}{n^5} \sum_{i=1}^n \sum_{j=1}^n \sum_{t=1}^n \sum_{l=1}^n \sum_{m=1}^n \frac{1}{h^{2d+2} f(\omega_j; \omega_i^c) f(\omega_l; \omega_i^c)} SK\left(\frac{\omega_t - \omega_j}{h}; \frac{\omega_t^c - \omega_i^c}{h}\right) SK\left(\frac{\omega_m - \omega_l}{h}; \frac{\omega_m^c - \omega_i^c}{h}\right) \epsilon_t \epsilon_m.
\end{aligned}$$

We show below with (i)-(iii) that  $T_{221} = \frac{1}{n h^{2+d-d_1}} B_2(1 + o_p(1)) + o_p(n^{-1} h^{-2-\frac{d}{2}})$ .

(i) When  $i \neq j \neq t \neq l \neq m$ , we apply Lemma 2 to perform U-statistics decomposition to have

$$\begin{aligned}
T_{221} &= \theta_n + \sum_{j=1}^5 \binom{5}{j} H_n^{(j)}(Q_{v_1}, \dots, Q_{v_j}). \text{ We can show that } \theta_n = 0, H_n^{(1)}(Q_{v_1}) = 0, H_n^{(2)}(Q_{v_1}, Q_{v_2}) = \\
&O_p(n^{-1} \sigma_{2n}) = O_p(n^{-1} h^{-2-\frac{d}{2}+\frac{d_1}{2}}) = o_p(n^{-1} h^{-2-\frac{d}{2}}) \text{ since } \sigma_{2n}^2 = O(h^{-4-(d-d_1)}), H_n^{(3)}(Q_{v_1}, Q_{v_2}, Q_{v_3}) = \\
&O_p(n^{-\frac{3}{2}} \sigma_{3n}) = O_p(n^{-\frac{3}{2}} (h^{-2-(d-d_1)} + h^{-2-\frac{d}{2}})) = o_p(n^{-1} h^{-2-\frac{d}{2}}) \text{ since } \sigma_{3n}^2 = O(h^{-4-(2d-2d_1)} + h^{-4-d}), \\
&H_n^{(4)}(Q_{v_1}, Q_{v_2}, Q_{v_3}, Q_{v_4}) = O_p(n^{-2} \sigma_{4n}) = O_p(n^{-2} (h^{-2-\frac{(d+d_1)}{2}} + h^{-2-d+\frac{d_1}{2}})) = o_p(n^{-1} h^{-2-\frac{d}{2}}) \text{ since } \\
&\sigma_{4n}^2 = O(h^{-4-(d+d_1)} + h^{-4-(2d-d_1)}), H_n^{(5)}(Q_{v_1}, Q_{v_2}, Q_{v_3}, Q_{v_4}, Q_{v_5}) = O_p(n^{-\frac{5}{2}} \sigma_{5n}) = O_p(n^{-\frac{5}{2}} (h^{-2-d} = \\
&o_p(n^{-1} h^{-2-\frac{d}{2}}) \text{ since } \sigma_{5n}^2 = O(h^{-4-2d}). \text{ Thus, } T_{221} = o_p(n^{-1} h^{-2-\frac{d}{2}}).
\end{aligned}$$

(ii) When  $i \neq j = l \neq t = m$ , or  $i \neq j = m \neq t = l$ , or  $i \neq j = l \neq t \neq m$ ,  $i \neq j = m \neq t \neq l$ ,  $i \neq j \neq t = t \neq m$ , we can show that  $T_{221} = o_p(n^{-1} h^{-2-\frac{d}{2}})$ .

(iii) When  $i \neq j \neq t = m \neq l$ , we can show that with similar arguments as in  $T_{211B}$  and  $T_{231B}$  that

$$\begin{aligned}
T_{221} &= \frac{1}{n^5} \sum_{i=1}^n \sum_{j=1}^n \sum_{t=1}^n \sum_{l=1}^n \frac{\epsilon_t^2}{h^{2d+2} f(\omega_j; \omega_i^c) f(\omega_l; \omega_i^c)} SK\left(\frac{\omega_t - \omega_j}{h}; \frac{\omega_t^c - \omega_i^c}{h}\right) SK\left(\frac{\omega_t - \omega_l}{h}; \frac{\omega_t^c - \omega_i^c}{h}\right) \\
&\quad t \neq j \neq i \neq l \\
&\equiv T_{221B} = \frac{1}{nh^{2+d-d_1}} B_2(1 + o_p(1)), \\
(b) \ T_{222} &= \frac{1}{n^5} \sum_{i=1}^n \sum_{j=1}^n \sum_{t=1}^n \sum_{l=1}^n \sum_{m=1}^n \frac{1}{h^{2d+2} f(\omega_j; \omega_i^c) f(\omega_l; \omega_i^c)} SK\left(\frac{\omega_t - \omega_j}{h}; \frac{\omega_t^c - \omega_i^c}{h}\right) SK\left(\frac{\omega_m - \omega_l}{h}; \frac{\omega_m^c - \omega_i^c}{h}\right) \epsilon_t \sum_{|\mathbf{k}|=p+1} h^{p+1} D^{\mathbf{k}} m_{ml;mi}. \\
&\quad t \neq j \neq i, m \neq l \neq i \\
T_{223} &= \frac{1}{n^5} \sum_{i=1}^n \sum_{j=1}^n \sum_{t=1}^n \sum_{l=1}^n \sum_{m=1}^n \frac{1}{h^{2d+2} f(\omega_j; \omega_i^c) f(\omega_l; \omega_i^c)} SK\left(\frac{\omega_t - \omega_j}{h}; \frac{\omega_t^c - \omega_i^c}{h}\right) SK\left(\frac{\omega_m - \omega_l}{h}; \frac{\omega_m^c - \omega_i^c}{h}\right) \epsilon_m \sum_{|\mathbf{k}|=p+1} h^{p+1} D^{\mathbf{k}} m_{tj;ti}. \\
&\quad t \neq j \neq i, m \neq l \neq i \\
T_{224} &= \frac{1}{n^5} \sum_{i=1}^n \sum_{j=1}^n \sum_{t=1}^n \sum_{l=1}^n \sum_{m=1}^n \frac{1}{h^{2d+2} f(\omega_j; \omega_i^c) f(\omega_l; \omega_i^c)} SK\left(\frac{\omega_t - \omega_j}{h}; \frac{\omega_t^c - \omega_i^c}{h}\right) SK\left(\frac{\omega_m - \omega_l}{h}; \frac{\omega_m^c - \omega_i^c}{h}\right) \\
&\quad t \neq j \neq i, m \neq l \neq i \\
&\quad \sum_{|\mathbf{k}|=p+1} h^{p+1} D^{\mathbf{k}} m_{tj;ti} \sum_{|\mathbf{k}|=p+1} h^{p+1} D^{\mathbf{k}} m_{ml;mi}.
\end{aligned}$$

We show similarly that  $T_{22i} = o_p(n^{-1}h^{-2-\frac{d}{2}})$  for  $i = 2, 3$ , and 4.

The claim in (3) follows from (3)(a) and (b) above and we conclude the proof for Theorem 1 based on (1)-(3). □

## Theorem 2.

*Proof.* Given that  $\hat{T} = T_{1n} + T_{2n} + T_{3n}$ , and Theorem 1 implies that  $T_{2n} = o_p(1)$ , we show below that

$$(1) \ T_{1n} = E[g_x(W_i) - E(g_x(\omega_l; \omega_i^c) | \omega_i^c)]^2 + o_p(1).$$

$$(2) \ T_{3n} = o_p(1).$$

Since under the alternative  $H_A : g_x(\omega; \omega^c)$  depends on  $\omega$ ,  $\hat{T} \rightarrow E[g_x(W_i) - E(g_x(\omega_l; \omega_i^c) | \omega_i^c)]^2 > 0$ . Then we have as  $n \rightarrow \infty$ ,  $P(nh^{2+\frac{d}{2}}\hat{T} > c_n) \rightarrow 1$  for any positive constant  $c_n = o(nh^{2+\frac{d}{2}})$ . Below we show (1) and (2).

$$(1) \ T_{1n} = \frac{1}{n} \sum_{i=1}^n g_x^2(\omega_i; \omega_i^c) - \frac{2}{n^2} \sum_{i=1}^n \sum_{l=1}^n g_x(\omega_i; \omega_i^c) g_x(\omega_l; \omega_l^c) + \frac{1}{n} \sum_{i=1}^n \left[ \frac{1}{n} \sum_{l=1}^n g_x(\omega_l; \omega_l^c) \right]^2 = T_{11} + T_{12} + T_{13}.$$

$$(a) \text{ With } E g_x^2(\omega_i; \omega_i^c) < C \text{ in A3(3), we easily have } T_{11} \xrightarrow{P} E g_x^2(\omega_i; \omega_i^c).$$

$$(b) \ T_{12} = -\frac{1}{n^2} \sum_{i=1}^n \sum_{l=1}^n [g_x(\omega_i; \omega_i^c) g_x(\omega_l; \omega_l^c) + g_x(\omega_l; \omega_l^c)] = -\frac{1}{n^2} \sum_{i=1}^n \sum_{l=1}^n \phi_{il}.$$

(i) When  $i \neq l$ , we apply Lemma 2 to obtain

$$T_{12} = -\frac{2}{n^2} \sum_{i=1}^n \sum_{l=1}^n \phi_{il} = -\left\{ \frac{2}{n} \sum_{i=1}^n [E(\phi_{il} | W_i) - E\phi_{il}] + E\phi_{il} + O_p(n^{-1}(E\phi_{il}^2)^{\frac{1}{2}}) \right\}.$$

Since  $E\phi_{il}^2 \leq C E g_x^2(\omega_i; \omega_i^c) g_x^2(\omega_l; \omega_l^c) \leq C [E g_x^4(\omega_i; \omega_i^c)]^{\frac{1}{2}} [E g_x^4(\omega_l; \omega_l^c)]^{\frac{1}{2}} < C$  by Cauchy Schwartz Inequality and A1(3),  $O_p(n^{-1}(E\phi_{il}^2)^{\frac{1}{2}}) = O_p(n^{-1})$ . Furthermore,  $\frac{2}{n} \sum_{i=1}^n [E(\phi_{il} | W_i) - E\phi_{il}] = O_p(n^{-\frac{1}{2}})$ , since  $E[E(\phi_{il} | W_i) - E\phi_{il}]^2 \leq C E\phi_{il}^2 = O(1)$ . Clearly,  $\phi_{il} = 2E g_x(\omega_i; \omega_i^c) g_x(\omega_l; \omega_l^c)$ , thus we obtain

$$T_{12} = -2E g_x(\omega_i; \omega_i^c) g_x(\omega_l; \omega_l^c) + o_p(1).$$

$$(ii) \text{ When } i = l, T_{12} = -\frac{2}{n^2} \sum_{i=1}^n g_x^2(\omega_i; \omega_i^c) = O_p(n^{-1}).$$

So in all,  $T_{12} = -2E g_x(\omega_i; \omega_i^c) g_x(\omega_l; \omega_l^c) + o_p(1)$ .

$$(c) \quad T_{13} = \frac{1}{n^3} \sum_{i=1}^n \sum_{l=1}^n \sum_{j=1}^n g_x(\omega_l; \omega_i^c) g_x(\omega_j; \omega_i^c).$$

(i) When  $i \neq j \neq l$ , we apply Lemma 2 to write it as a U-statistics, perform U-Statistics decomposition,

analyze each term and obtain

$$T_{13} = E g_x(\omega_l; \omega_i^c) g_x(\omega_j; \omega_i^c) + o_p(1).$$

(ii) When  $i = l \neq j$ , or  $i = j \neq l$ , or  $j = l \neq i$ , or  $i = l = j$ ,  $T_{13} = O(n^{-1})$  since  $E|T_{13}| = O(n^{-1})$ .

So in all, we obtain  $T_{13} = E g_x(\omega_l; \omega_i^c) g_x(\omega_j; \omega_i^c) + o_p(1)$ .

Combining (1)(a)-(c) above, we obtain

$$\begin{aligned} T_{1n} &= E g_x^2(\omega_i; \omega_i^c) - 2E g_x(\omega_i; \omega_i^c) g_x(\omega_l; \omega_i^c) + E g_x(\omega_l; \omega_i^c) g_x(\omega_j; \omega_i^c) + o_p(1) \\ &= E [g_x(W_i) - E(g_x(\omega_l; \omega_i^c) | \omega_i^c)]^2 + o_p(1). \end{aligned}$$

$$\begin{aligned} (2) \quad T_{3n} &= \frac{2}{n} \sum_{i=1}^n [g_x(W_i) - \frac{1}{n} \sum_{\substack{l=1 \\ l \neq i}}^n g_x(\omega_l; \omega_i^c)] [\hat{g}_x(W_i) - g_x(W_i)] \\ &\quad - \frac{2}{n} \sum_{i=1}^n [g_x(W_i) - \frac{1}{n} \sum_{\substack{l=1 \\ l \neq i}}^n g_x(\omega_l; \omega_i^c)] \frac{1}{n} \sum_{l=1}^n [\hat{g}_x(\omega_l; \omega_i^c) - g_x(\omega_l; \omega_i^c)] \\ &= T_{31} + T_{32}. \end{aligned}$$

Given that  $\frac{1}{n} \sum_{\substack{l=1 \\ l \neq i}}^n g_x(\omega_l; \omega_i^c) = E(g_x(\omega_l; \omega_i^c) | \omega_i^c) + o_p(1)$  uniformly for  $\omega_i^c$ , we denote  $D^{\mathbf{k}} m_{ji} \equiv (D^{\mathbf{k}} m)(W_i + \lambda(W_j - W_i))(\frac{W_j - W_i}{h})^{\mathbf{k}}$ . Then

$$\begin{aligned} (a) \quad T_{31} &= \frac{2}{n} \sum_{i=1}^n [g_x(W_i) - E g_x(\omega_l; \omega_i^c) | \omega_i^c] \left[ \frac{1}{f(W_i) n h^{d+1}} \sum_{\substack{j=1 \\ j \neq i}}^n SK\left(\frac{W_j - W_i}{h}\right) (\epsilon_j + \sum_{0 \leq |\mathbf{k}| \leq p} \frac{h^{p+1}}{\mathbf{k}!} D^{\mathbf{k}} m_{ji}) (1 + o_p(1)) \right] \\ &= (T_{311} + T_{312})(1 + o_p(1)). \end{aligned}$$

$$(i) \quad T_{311} = \frac{2}{n^2} \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n [g_x(W_i) - E g_x(\omega_l; \omega_i^c) | \omega_i^c] \frac{1}{f(W_i) n h^{d+1}} SK\left(\frac{W_j - W_i}{h}\right) \epsilon_j = \frac{2}{n^2} \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \psi_{nij}.$$

We apply Lemma 2 to write  $\phi_{nij} = \psi_{nij} + \psi_{nji}$ , and

$$T_{311} = \binom{n}{2}^{-1} \sum_{i=1}^n \sum_{\substack{j=1 \\ i < j}}^n \phi_{nij} (1 + o_p(1)) = \theta_n + \sum_{j=1}^2 \binom{2}{j} H_n^{(j)}(Q_{v_1}, \dots, Q_{v_j}).$$

$\theta_n = 0$ ,  $H_n^{(1)}(Q_{v_1}) = O_p(n^{-\frac{1}{2}} \sigma_{1n}) = O_p(n^{-\frac{1}{2}} h)$  since  $\sigma_{1n}^2 = O(h^2)$ ,  $H_n^{(2)}(Q_{v_1}, Q_{v_2}) = O_p(n^{-1} \sigma_{2n}) = O_p(n^{-1} h^{-1-\frac{d}{2}})$  since  $\sigma_{2n}^2 = O(h^{-2-d})$ . Thus,  $T_{311} = O_p(n^{-\frac{1}{2}} h)$ .

(ii)  $T_{312} = \frac{2}{n^2} \sum_{i=1}^n \sum_{\substack{j=1 \\ i \neq j}}^n [g_x(W_i) - E g_x(\omega_l; \omega_i^c) | \omega_i^c] \frac{1}{f(W_i) n h^{d+1}} SK\left(\frac{W_j - W_i}{h}\right) \sum_{0 \leq |\mathbf{k}| \leq p} \frac{h^{p+1}}{\mathbf{k}!} D^{\mathbf{k}} m_{ji} (1 + o_p(1)) = O_p(h^p)$ , since  $E|T_{312}| = O(h^p)$ .

So we obtain  $T_{31} = O_p(n^{-\frac{1}{2}} h) + O_p(h^p) = o_p(1)$ .

(b) With similar arguments, we show that

$$\begin{aligned} T_{32} &= \frac{2}{n} \sum_{i=1}^n [g_x(W_i) - E g_x(\omega_l; \omega_i^c) | \omega_i^c] \frac{1}{n} \sum_{l=1}^n \frac{1}{f(\omega_l; \omega_i^c) n h^{d+1}} \sum_{t=1}^n SK\left(\frac{\omega_t - \omega_l}{h}; \frac{\omega_t^c - \omega_l^c}{h}\right) [\epsilon_t \\ &\quad + \sum_{0 \leq |\mathbf{k}| \leq p} \frac{h^{p+1}}{\mathbf{k}!} D^{\mathbf{k}} m_{tl; ti}] (1 + o_p(1)) \\ &= o_p(1). \end{aligned}$$

The claim in (2) follows from (2)(a)-(b) above.

Thus, we obtain  $T_{3n} = o_p(1)$  and we conclude the proof of Theorem 2.

□

### Theorem 3.

*Proof.* Following the arguments in Lemma 1, we define  $t_{n,j}^*(W) = \frac{1}{nh^d} \sum_{i=1}^n K(\frac{W_i - W}{h})(\frac{W_i - W}{h})^j \epsilon_j^*$ ,  $(\tau_{n,|j|}^*)_k =$

$$t_{n,G_{|j|}(k)}^*, \text{ and } \tau_n^*(W) = \begin{bmatrix} \tau_{n,0}^*(W) \\ \tau_{n,1}^*(W) \\ \vdots \\ \tau_{n,p}^*(W) \end{bmatrix}, \text{ then we obtain from the first order condition in equation (1.2),}$$

$$h\hat{g}_x^*(W) = e'_{N,1+d} S_n(W)^{-1} \tau_n^*(W) = \frac{1}{f(W)} [S^{-1}]_{1+d, \cdot} \tau_n^*(W) (1 + o_p(1)).$$

From these, we obtain

$$\hat{g}_x^*(W_i) = \frac{1}{f(W_i)nh^{d+1}} \sum_{\substack{j=1 \\ j \neq i}}^n SK(\frac{W_j - W_i}{h}) \epsilon_j^* (1 + o_p(1)), \text{ and}$$

$$\hat{g}_x^*(\omega_l; \omega_i^c) = \frac{1}{f(\omega_l; \omega_i^c)nh^{d+1}} \sum_{\substack{j=1 \\ j \neq l, j \neq i}}^n SK(\frac{\omega_j - \omega_l}{h}; \frac{\omega_j^c - \omega_i^c}{h}) \epsilon_j^* (1 + o_p(1)).$$

$$\hat{T}^* = \frac{1}{n} \sum_{i=1}^n (\hat{g}_x^*(W_i))^2 + \frac{1}{n} \sum_{i=1}^n [\frac{1}{n} \sum_{\substack{l=1 \\ l \neq i}}^n \hat{g}_x^*(\omega_l; \omega_i^c)]^2 - \frac{2}{n^2} \sum_{i=1}^n \sum_{\substack{l=1 \\ l \neq i}}^n \hat{g}_x^*(W_i) \hat{g}_x^*(\omega_l; \omega_i^c) = T_1^* + T_2^* + T_3^*.$$

We show below that conditional on  $Q_{(n)} = \{W_i, y_i\}_{i=1}^n$ ,

$$(1) \quad nh^{2+\frac{d}{2}} (T_1^* - \frac{1}{nh^{2+d}} B_1 (1 + o_p(1))) \xrightarrow{d} N(0, \Omega).$$

$$(2) \quad T_3^* = -\frac{2}{nh^{2+d-d_1}} B_3 (1 + o_p(1)) + o_p((nh^{2+d})^{-1}).$$

$$(3) \quad T_2^* = \frac{1}{nh^{2+d-d_1}} B_2 (1 + o_p(1)) + o_p((nh^{2+d})^{-1}).$$

The claim in Theorem 3 follows from (1)-(3).

$$\begin{aligned} (1) \quad T_1^* &= \frac{1}{n} \sum_{i=1}^n [\frac{1}{f(W_i)nh^{d+1}} \sum_{\substack{j=1 \\ j \neq i}}^n SK(\frac{W_j - W_i}{h}) \epsilon_j^*]^2 \\ &= \frac{1}{n^3} \sum_{i=1}^n \sum_{j=1}^n \sum_{\substack{l=1 \\ l \neq i, j}}^n \underbrace{\frac{1}{f^2(W_i)nh^{2(d+1)}} SK(\frac{W_j - W_i}{h}) SK(\frac{W_l - W_i}{h}) \epsilon_j^* \epsilon_l^*}_{\psi_{nijl}} (1 + o_p(1)). \end{aligned}$$

(a) When  $i \neq j \neq l$ , we note that the U-statistics results in Lemma 2 can not be applied here as we do not have the IID assumption conditioning on the data  $Q_{(n)}$ . We let  $\phi_{nijl} = \psi_{nijl} + \psi_{njil} + \psi_{nlji}$ , which is symmetric in  $i, j, l$ . We rewrite  $T_1^*$  as

$$\begin{aligned}
T_1^* &= (1 + o_p(1)) \left\{ \frac{1}{3} \binom{n}{3}^{-1} \binom{n-2}{1} \sum_{i=1}^n \sum_{j=1}^n \phi_{2nij} + \frac{1}{3} \binom{n}{3}^{-1} \left[ \sum_{i=1}^n \sum_{j=1}^n \sum_{l=1}^n \phi_{nijl} - \binom{n-2}{1} \sum_{i=1}^n \sum_{j=1}^n \phi_{2nij} \right] \right\} \\
&= (1 + o_p(1)) \{T_{1a}^* + T_{1b}^*\},
\end{aligned}$$

where  $\phi_{2nij} = \int \psi_{nijl} F(W_l) dW_l = \frac{\epsilon_j^* \epsilon_i^*}{h^{2(d+1)}} \int \frac{1}{f^2(W_l)} SK\left(\frac{W_j - W_l}{h}\right) SK\left(\frac{W_i - W_l}{h}\right) f(W_l) dW_l$ .

(i) Claim:  $T_{1b}^* = o_p(n^{-1}h^{-2-\frac{d}{2}})$ .

$$T_{1b}^* = \frac{1}{3} \binom{n}{3}^{-1} \left[ \sum_{i=1}^n \sum_{j=1}^n \sum_{l=1}^n \phi_{nijl} - \phi_{2nij} - \phi_{2nil} - \phi_{2njl} \right] = \frac{1}{3} \binom{n}{3}^{-1} \sum_{i=1}^n \sum_{j=1}^n \sum_{l=1}^n \Phi_{nijl}.$$

Note that  $E^*(\epsilon_i^*) \equiv E(\epsilon_i^* | Q_{(n)}) = 0$ , so  $E(T_{1b}^* | Q_{(n)}) = 0$ .

$$\begin{aligned}
V(T_{1b}^* | Q_{(n)}) &= \frac{1}{9} \binom{n}{3}^{-2} \sum_{i=1}^n \sum_{j=1}^n \sum_{l=1}^n \sum_{i'=1}^n \sum_{j'=1}^n \sum_{l'=1}^n E(\Phi_{nijl} \Phi_{ni'j'l'} | Q_{(n)}) \\
&= \frac{1}{9} \binom{n}{3}^{-2} \sum_{i=1}^n \sum_{j=1}^n \sum_{l=1}^n E(\Phi_{nijl}^2 | Q_{(n)}) + V_{1b1}.
\end{aligned}$$

Note that if  $(i, j, l)$  are each distinct from  $(i', j', l')$ , since  $\epsilon_i^*$  is independent conditioning on  $Q_{(n)}$ , then  $E(\Phi_{nijl} \Phi_{ni'j'l'} | Q_{(n)}) = E(\Phi_{nijl} | Q_{(n)}) E(\Phi_{ni'j'l'} | Q_{(n)}) = 0$ . Similarly, if only one index in  $(i, j, l)$  is the same as that in  $(i', j', l')$ ,  $E(\Phi_{nijl} \Phi_{ni'j'l'} | Q_{(n)}) = 0$ . So in  $V_{1b1}$ , we only consider the case that two of the indices in  $(i, j, l)$  are the same as that in  $(i', j', l')$ . Due to the symmetry in  $\Phi_{nijl}$ , we can just consider any two indices.

Consider  $i = i' \neq j = j' \neq l \neq l'$ ,  $E(\Phi_{nijl} \Phi_{nijl'} | Q_{(n)}) = E[(\psi_{nlji} - \phi_{2nij})(\psi_{nl'ji} - \phi_{2nij}) | Q_{(n)}]$ . Thus,

$$\begin{aligned}
V_{1b1} &= O\left(\frac{1}{9} \binom{n}{3}^{-2} \sum_{i=1}^n \sum_{j=1}^n \sum_{l=1}^n \sum_{i'=1}^n \sum_{j'=1}^n \sum_{l'=1}^n E[(\psi_{nlji} - \phi_{2nij})(\psi_{nl'ji} - \phi_{2nij}) | Q_{(n)}]\right) \\
&= O\left(\frac{1}{9} \binom{n}{3}^{-2} \sum_{i=1}^n \sum_{j=1}^n \sum_{l=1}^n \sum_{i'=1}^n \sum_{j'=1}^n \sum_{l'=1}^n \frac{\epsilon_i^2 \epsilon_j^2}{h^{4(d+1)}} \left[ \frac{1}{f^2(W_l)} SK\left(\frac{W_j - W_l}{h}\right) SK\left(\frac{W_i - W_l}{h}\right) \right. \right. \\
&\quad \left. \left. - \int \frac{1}{f^2(W_l)} SK\left(\frac{W_j - W_l}{h}\right) SK\left(\frac{W_i - W_l}{h}\right) f(W_l) dW_l \right] \right. \\
&\quad \left. \times \left[ \frac{1}{f^2(W_{l'})} SK\left(\frac{W_j - W_{l'}}{h}\right) SK\left(\frac{W_i - W_{l'}}{h}\right) - \int \frac{1}{f^2(W_{l'})} SK\left(\frac{W_j - W_{l'}}{h}\right) SK\left(\frac{W_i - W_{l'}}{h}\right) f(W_{l'}) dW_{l'} \right] \right) \\
&= O\left(\frac{1}{9} \binom{n}{3}^{-2} \sum_{i=1}^n \sum_{j=1}^n \sum_{l=1}^n \sum_{i'=1}^n \sum_{j'=1}^n \sum_{l'=1}^n \frac{\epsilon_i^2 \epsilon_j^2}{h^{4(d+1)}} \left[ \frac{1}{f^2(W_l)} SK\left(\frac{W_j - W_l}{h}\right) SK\left(\frac{W_i - W_l}{h}\right) \right. \right. \\
&\quad \left. \left. - \int \frac{1}{f^2(W_l)} SK\left(\frac{W_j - W_l}{h}\right) SK\left(\frac{W_i - W_l}{h}\right) f(W_l) dW_l \right] \right. \\
&\quad \left. \times \left[ \frac{1}{f^2(W_{l'})} SK\left(\frac{W_j - W_{l'}}{h}\right) SK\left(\frac{W_i - W_{l'}}{h}\right) - \int \frac{1}{f^2(W_{l'})} SK\left(\frac{W_j - W_{l'}}{h}\right) SK\left(\frac{W_i - W_{l'}}{h}\right) f(W_{l'}) dW_{l'} \right] (1 + o_p(1)) \right) \\
&= O\left(\frac{1}{9} \binom{n}{3}^{-2} \sum_{i=1}^n \sum_{j=1}^n \sum_{l=1}^n \sum_{i'=1}^n \sum_{j'=1}^n \sum_{l'=1}^n \psi_{nvijll'} (1 + o_p(1))\right) = O(V_{1b11}),
\end{aligned}$$

where the third to last equality is from the fact that  $\hat{\epsilon}_i = y_i - \hat{m}(W_i) = \epsilon_i + (m(W_i) - \hat{m}(W_i)) = \epsilon_i + o_p(1)$

uniformly  $\forall W_i \in G$ .

The claim that  $\sup_{W \in G} |\hat{m}(W) - m(W)| = o_p(1)$  follows by applying Lemma 3 of Martins-Filho et al. (2018). Note that assumption A3(1) implies that  $E|\epsilon|^s < C$  for some  $s > 2$ . This observation, together with assumptions A1(1)-(3), A2, A3(1), (2) and A4(1), enable us to apply Lemma 3 of Martins-Filho et al. (2018) to obtain the uniform convergence result. We apply Lemma 2 on  $V_{1b11}$  below. Define

$$\begin{aligned} \phi_{nvijll'} = & \psi_{nvijll'} + \psi_{nvijl'l} + \psi_{nviljl'} + \psi_{nvill'j} + \psi_{nvil'jl} + \psi_{nvil'lj} \\ & + \psi_{nvjill'} + \psi_{nvjil'l} + \psi_{nvjlil'} + \psi_{nvjll'i} + \psi_{nvjl'il} + \psi_{nvjl'li} \\ & + \psi_{nvlijl'} + \psi_{nvli'lj} + \psi_{nvlij'l} + \psi_{nvlijli} + \psi_{nvll'ij} + \psi_{nvll'ji} \\ & + \psi_{nvll'ijl} + \psi_{nvll'ilj} + \psi_{nvll'jil} + \psi_{nvll'jli} + \psi_{nvll'lij} + \psi_{nvll'lji}, \\ V_{1b11} = & \frac{1}{9} \binom{n}{3}^{-2} \binom{n}{4} \binom{n}{4}^{-1} \sum_{i=1}^n \sum_{j=1}^n \sum_{l=1}^n \sum_{i < j < l}^n \phi_{nvijll'} = O(n^{-2}) [\theta_n + \sum_{j=1}^4 \binom{4}{j} H_n^{(j)}(Q_{v_1}, \dots, Q_{v_j})]. \end{aligned}$$

Since  $\int [\frac{1}{f^2(W_l)} SK(\frac{W_j - W_l}{h}) SK(\frac{W_i - W_l}{h}) - \int \frac{1}{f^2(W_l)} SK(\frac{W_j - W_l}{h}) SK(\frac{W_i - W_l}{h}) f(W_l) dW_l] f(W_l) dW_l = 0$ ,  $\theta_n = 0$ , and  $H_n^{(1)}(Q_{v_1}) = 0$ .  $H_n^{(2)}(Q_{v_1}, Q_{v_2}) = O_p(n^{-1} \sigma_{2n}) = O_p(n^{-1} h^{-4 - \frac{3d}{2}})$ , since  $\sigma_{2n}^2 = E[E^2(\phi_{nvijll'} | Q_l, Q_{l'})] = O(h^{-8-3d})$ .  $H_n^{(3)}(Q_{v_1}, Q_{v_2}, Q_{v_3}) = O_p^2(n^{-\frac{3}{2}} \sigma_{3n}) = O_p(n^{-\frac{3}{2}} h^{-4-2d})$ , since  $\sigma_{3n}^2 = E[E^2(\phi_{nvijll'} | Q_i, Q_j, Q_l)] = O(h^{-8-4d})$ .  $H_n^{(4)}(Q_{v_1}, Q_{v_2}, Q_{v_3}, Q_{v_4}) = O_p(n^{-2} \sigma_{4n}) = O_p(n^{-2} h^{-4 - \frac{5}{2}d})$ , since  $\sigma_{4n}^2 = E[\phi_{nvijll'}^2] = O(h^{-8-5d})$ .

So  $V_{1b11} = O_p(n^{-2} [n^{-1} h^{-4 - \frac{3d}{2}} + n^{-\frac{3}{2}} h^{-4-2d} + n^{-2} h^{-4 - \frac{5}{2}d}]) = O_p(n^{-3} h^{-4 - \frac{3}{2}d})$ , and  $V_{1b1} = O_p(n^{-3} h^{-4 - \frac{3}{2}d})$ .

Next, we note that by  $c$ - $r$  inequality, and since  $E(\phi_{2nij}^2 | Q_{(n)})$  is of smaller order,

$$\begin{aligned} & \binom{n}{3}^{-2} \sum_{i=1}^n \sum_{j=1}^n \sum_{l=1}^n E(\Phi_{nijl}^2 | Q_{(n)}) \\ & \leq C \binom{n}{3}^{-2} \sum_{i=1}^n \sum_{j=1}^n \sum_{l=1}^n [E(\phi_{nijl}^2 | Q_{(n)}) + E(\phi_{2nij}^2 | Q_{(n)}) + E(\phi_{2nil}^2 | Q_{(n)}) + E(\phi_{2njl}^2 | Q_{(n)})] \\ & = C \binom{n}{3}^{-2} \sum_{i=1}^n \sum_{j=1}^n \sum_{l=1}^n E(\phi_{nijl}^2 | Q_{(n)}) (1 + o_p(1)) \\ & = C \binom{n}{3}^{-2} \sum_{i=1}^n \sum_{j=1}^n \sum_{l=1}^n [E(\psi_{nijl}^2 | Q_{(n)}) + E(\psi_{njil}^2 | Q_{(n)}) + E(\psi_{nlji}^2 | Q_{(n)})] (1 + o_p(1)) \\ & = O_p(n^{-3} h^{-4-2d}), \end{aligned}$$

where the last equality follows from the observation that

$$\begin{aligned}
& \binom{n}{3}^{-2} \sum_{i=1}^n \sum_{j=1}^n \sum_{l=1}^n E(\psi_{nijl}^2 | Q_{(n)}) \\
&= \binom{n}{3}^{-2} \sum_{i=1}^n \sum_{j=1}^n \sum_{l=1}^n \frac{\epsilon_j^2 \epsilon_l^2}{f^4(W_i) h^{4(d+1)}} SK^2\left(\frac{W_j - W_i}{h}\right) SK^2\left(\frac{W_l - W_i}{h}\right) \\
&= \binom{n}{3}^{-2} \sum_{i=1}^n \sum_{j=1}^n \sum_{l=1}^n \frac{\epsilon_j^2 \epsilon_l^2}{f^4(W_i) h^{4(d+1)}} SK^2\left(\frac{W_j - W_i}{h}\right) SK^2\left(\frac{W_l - W_i}{h}\right) (1 + o_p(1)) \\
&= O_p(n^{-3} h^{-4-2d}).
\end{aligned}$$

$$\text{So } V(T_{1b}^* | Q_{(n)}) = \frac{1}{9} \binom{n}{3}^{-2} \sum_{i=1}^n \sum_{j=1}^n \sum_{l=1}^n E(\Phi_{nijl}^2 | Q_{(n)}) + V_{1b1} = O_p(n^{-3} h^{-4-2d}) + O_p(n^{-3} h^{-4-\frac{3}{2}d}).$$

Thus,  $T_{1b}^* = O_p(n^{-\frac{3}{2}} h^{-2-d}) = o_p(n^{-1} h^{-2-\frac{d}{2}})$ , as claimed.

(ii) Claim:  $nh^{2+\frac{d}{2}} T_{1a}^* \xrightarrow{d} N(0, \Omega)$ .

$$\begin{aligned}
nh^{2+\frac{d}{2}} T_{1a}^* &= nh^{2+\frac{d}{2}} \frac{1}{3} \binom{n}{3}^{-1} \binom{n-2}{1} \sum_{i=1}^n \sum_{j=1}^n \phi_{2nij} \\
&= \sum_{i=1}^n \sum_{j=1}^n \sum_{i < j} nh^{2+\frac{d}{2}} \binom{n}{2}^{-1} \frac{\epsilon_i^* \epsilon_j^*}{h^{2(d+1)}} \int \frac{1}{f^2(W_l)} SK\left(\frac{W_j - W_i}{h}\right) SK\left(\frac{W_l - W_i}{h}\right) f(W_l) dW_l \\
&= \sum_{i=1}^n \sum_{j=1}^n \sum_{i < j} \psi_{nij}.
\end{aligned}$$

Note that  $\psi_{nij}$  is symmetric in  $i, j$ ,  $E(\psi_{nij} | Q_{(n)}, \epsilon_j^*) = 0$ , so conditioning on  $Q_{(n)}$ ,  $nh^{2+\frac{d}{2}} T_{1a}^*$  is a degenerate second order U-statistics. Defining  $(S_n^*)^2 \equiv E\left(\left(\sum_{i=1}^n \sum_{j=1}^n \sum_{i < j} \psi_{nij}\right)^2 | Q_{(n)}\right) = V(T_{1a}^* | Q_{(n)})$ , we apply

Proposition 3.2 of de Jong (1987) to obtain

$$(S_n^*)^{-1} nh^{2+\frac{d}{2}} T_{1a}^* \xrightarrow{d} N(0, 1),$$

if  $G_I, G_{II}, G_{IV}$  are each of order  $o_p((S_n^*)^4)$ , where  $G_I = \sum_{i=1}^n \sum_{j=1}^n \sum_{i < j} E(\psi_{nij}^4 | Q_{(n)})$ ,

$$G_{II} = \sum_{i=1}^n \sum_{j=1}^n \sum_{t=1}^n \sum_{i < j < t} [E(\psi_{nij}^2 \psi_{nit}^2 | Q_{(n)}) + E(\psi_{nji}^2 \psi_{njt}^2 | Q_{(n)}) + E(\psi_{niti}^2 \psi_{niti}^2 | Q_{(n)})] = G_{II1} + G_{II2} + G_{II3},$$

$$\begin{aligned}
G_{IV} &= \sum_{i=1}^n \sum_{j=1}^n \sum_{t=1}^n \sum_{i < j < t < l} [E(\psi_{nij} \psi_{nit} \psi_{nlt} \psi_{nlt} | Q_{(n)}) + E(\psi_{nij} \psi_{nil} \psi_{nlt} \psi_{nlt} | Q_{(n)}) + E(\psi_{nit} \psi_{nil} \psi_{nlt} \psi_{nlt} | Q_{(n)})] \\
&= G_{IV1} + G_{IV2} + G_{IV3}.
\end{aligned}$$

Since  $E(\epsilon_i^* \epsilon_j^* \epsilon_t^* \epsilon_m^* | Q_{(n)}) = \hat{\epsilon}_i^2 \hat{\epsilon}_j^2$  when  $i = t < j = m$ , and zero in the other cases, we have



$$\begin{aligned}
(S_n^*)^2 &= n^2 h^{4+d} \binom{n}{2}^{-2} \sum_{i=1}^n \sum_{j=1}^n \frac{\hat{\epsilon}_i^2 \hat{\epsilon}_j^2}{h^{4(d+1)}} [\int \frac{1}{f^2(W_l)} SK(\frac{W_j - W_l}{h}) SK(\frac{W_i - W_l}{h}) f(W_l) dW_l]^2 \\
&= \frac{4}{(n-1)^2} \sum_{i=1}^n \sum_{j=1}^n \frac{\epsilon_i^2 \epsilon_j^2}{h^{4(d+1)}} [\int \frac{1}{f^2(W_l)} SK(\frac{W_j - W_l}{h}) SK(\frac{W_i - W_l}{h}) f(W_l) dW_l]^2 (1 + o_p(1)) \\
&= (1 + o_p(1)) 2 \binom{n}{2}^{-1} \sum_{i=1}^n \sum_{j=1}^n \frac{\epsilon_i^2 \epsilon_j^2}{h^{4(d+1)}} [\int \frac{1}{f^2(W_l)} SK(\frac{W_j - W_l}{h}) SK(\frac{W_i - W_l}{h}) f(W_l) dW_l]^2 \\
&= (1 + o_p(1)) 2 \binom{n}{2}^{-1} \sum_{i=1}^n \sum_{j=1}^n \psi_{nsij},
\end{aligned}$$

where the second equality follows since  $n^2 \binom{n}{2}^{-2} = \frac{4}{(n-1)^2}$ , and  $\hat{\epsilon}_i = \epsilon_i + o_p(1)$  uniformly. The third

follows since  $\frac{2}{(n-1)^2} - \binom{n}{2}^{-1} = O(n^{-3})$ . We then apply Lemma 2 to obtain that

$$\binom{n}{2}^{-1} \sum_{i=1}^n \sum_{j=1}^n \psi_{nsij} = \theta_n + \sum_{j=1}^2 \binom{2}{j} H_n^{(j)}(Q_{v_1}, \dots, Q_{v_j}) = \frac{\Omega}{2} + O_p(n^{-1} h^{-\frac{d}{2}}),$$

since  $\theta_n = E\psi_{nsij} \rightarrow \int \sigma^4(W_i) dW_i \int [\int SK(\Psi_1 + \Psi) SK(\Psi) d\Psi]^2 d\Psi_1 = \frac{\Omega}{2}$ ,  $H_n^{(1)}(Q_{v_1}) = O_p(n^{-1})$ , and  $H_n^{(2)}(Q_{v_1}, Q_{v_2}) = O_p(n^{-1} h^{-\frac{d}{2}})$ . Thus,  $(S_n^*)^2 = \Omega + o_p(1)$ .

$G_I = \sum_{i=1}^n \sum_{j=1}^n h^{8+2d} \frac{16}{(n-1)^4} \frac{E((\epsilon_i^* \epsilon_j^*)^4 | Q_{(n)})}{h^{8(d+1)}} [\int \frac{1}{f^2(W_l)} SK(\frac{W_j - W_l}{h}) SK(\frac{W_i - W_l}{h}) f(W_l) dW_l]^4 = O(n^{-2} h^{-d}) = o_p((S_n^*)^4)$ , since  $(S_n^*)^4 = O(1)$ .

$$\begin{aligned}
G_{II1} &= \sum_{i=1}^n \sum_{j=1}^n \sum_{t=1}^n h^{8+2d} \frac{16}{(n-1)^4} h^{-8-8d} E((\epsilon_i^*)^4 (\epsilon_j^*)^2 (\epsilon_t^*)^2 | Q_{(n)}) [\int \frac{1}{f^2(W_l)} SK(\frac{W_j - W_l}{h}) SK(\frac{W_i - W_l}{h}) f(W_l) dW_l]^2 \\
&\quad \times [\int \frac{1}{f^2(W_{l'})} SK(\frac{W_t - W_{l'}}{h}) SK(\frac{W_i - W_{l'}}{h}) f(W_{l'}) dW_{l'}]^2 \\
&= O(n^{-1}).
\end{aligned}$$

Similarly,  $G_{II2} = O(n^{-1})$ ,  $G_{II3} = O(n^{-1})$ , thus we conclude that  $G_{II} = O(n^{-1}) = o((S_n^*)^4)$ .

$$\begin{aligned}
G_{IV1} &= \sum_{i=1}^n \sum_{j=1}^n \sum_{t=1}^n \sum_{l=1}^n \frac{16}{(n-1)^4} h^{-2d} E((\epsilon_i^*)^2 (\epsilon_j^*)^2 (\epsilon_t^*)^2 (\epsilon_l^*)^2 | Q_{(n)}) [\frac{1}{h^d} \int \frac{1}{f^2(W_l)} SK(\frac{W_j - W_l}{h}) SK(\frac{W_i - W_l}{h}) f(W_l) dW_l] \\
&\quad \times [\frac{1}{h^d} \int \frac{1}{f^2(W_l)} SK(\frac{W_t - W_l}{h}) SK(\frac{W_i - W_l}{h}) f(W_l) dW_l] [\frac{1}{h^d} \int \frac{1}{f^2(W_{l'})} SK(\frac{W_j - W_{l'}}{h}) SK(\frac{W_i - W_{l'}}{h}) f(W_{l'}) dW_{l'}] \\
&\quad \times [\frac{1}{h^d} \int \frac{1}{f^2(W_{l'})} SK(\frac{W_t - W_{l'}}{h}) SK(\frac{W_i - W_{l'}}{h}) f(W_{l'}) dW_{l'}] (1 + o_p(1)) \\
&= O(h^d).
\end{aligned}$$

Similarly,  $G_{IV2} = O(h^d)$ , and  $G_{IV3} = O(h^d)$ , and thus we have  $G_{IV} = O(h^d) = o((S_n^*)^4)$ .

Since we obtain that  $(S_n^*)^2 \xrightarrow{p} \Omega$ , we conclude that  $nh^{2+\frac{d}{2}} T_{1a}^* \xrightarrow{d} N(0, \Omega)$ .

(b) When  $i \neq j = l$ ,  $T_1^* \equiv T_{1B}^* = \frac{1}{nh^{2+d}} \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \frac{1}{f^2(W_i) h^d} SK^2(\frac{W_j - W_i}{h}) (\epsilon_j^*)^2 = \frac{1}{nh^{2+d}} T_{1B1}^*$ .

Claim:  $T_{1B}^* = \frac{1}{nh^{2+d}} B_1(1 + o_p(1))$ .

To this end, we only need to show that  $T_{1B1}^* = B_1 + o_p(1)$  conditioning on  $Q_{(n)}$ .

$$(i) \quad E(T_{1B1}^* | Q_{(n)}) = \frac{1}{n^2} \sum_{i=1}^n \sum_{\substack{j=1 \\ i \neq j}}^n \frac{1}{f^2(W_i)h^d} SK^2\left(\frac{W_j - W_i}{h}\right) \hat{\epsilon}_j^2 = B_1(1 + o_p(1)).$$

$$\begin{aligned} (ii) \quad V(T_{1B1}^* | Q_{(n)}) &= E([T_{1B1}^* - E(T_{1B1}^* | Q_{(n)})]^2 | Q_{(n)}) \\ &= E\left(\left[\frac{1}{n^2} \sum_{i=1}^n \sum_{\substack{j=1 \\ i \neq j}}^n \frac{1}{f^2(W_i)h^d} SK^2\left(\frac{W_j - W_i}{h}\right) ((\epsilon_j^*)^2 - \hat{\epsilon}_j^2)\right]^2 \middle| Q_{(n)}\right) \\ &= \frac{1}{n^4} \sum_{i=1}^n \sum_{\substack{j=1 \\ i \neq j}}^n \sum_{l=1}^n \sum_{\substack{m=1 \\ l \neq m}}^n \frac{1}{f^2(W_i)f^2(W_l)h^{2d}} SK^2\left(\frac{W_j - W_i}{h}\right) SK^2\left(\frac{W_m - W_l}{h}\right) E[((\epsilon_j^*)^2 - \hat{\epsilon}_j^2)((\epsilon_m^*)^2 - \hat{\epsilon}_m^2) | Q_{(n)}]. \end{aligned}$$

Note first that whenever  $j \neq m$ , above is zero. Second, the two point distribution  $\hat{F}_i$  described in Step 1 of the bootstrap procedure implies that  $E(\epsilon_j^* | Q_{(n)}) = 0$ ,  $E(|\epsilon_j^*| | Q_{(n)}) = \frac{2}{\sqrt{5}}|\hat{\epsilon}_j|$ ,  $E((\epsilon_j^*)^2 | Q_{(n)}) = \hat{\epsilon}_j^2$ ,  $E((\epsilon_j^*)^3 | Q_{(n)}) = 0$ ,  $E((\epsilon_j^*)^4 | Q_{(n)}) = 2\hat{\epsilon}_j^4$ .

When  $i \neq j = m \neq l$ , as  $E(((\epsilon_j^*)^2 - \hat{\epsilon}_j^2)^2 | Q_{(n)}) = \hat{\epsilon}_j^4$ ,

$$\begin{aligned} &V(T_{1B1}^* | Q_{(n)}) \\ &= \frac{1}{n^4} \sum_{i=1}^n \sum_{\substack{j=1 \\ i \neq j}}^n \sum_{\substack{l=1 \\ i \neq j \neq l}}^n \frac{1}{f^2(W_i)f^2(W_l)h^{2d}} SK^2\left(\frac{W_j - W_i}{h}\right) SK^2\left(\frac{W_j - W_l}{h}\right) E(((\epsilon_j^*)^2 - \hat{\epsilon}_j^2)^2 | Q_{(n)}) \\ &= \frac{1}{n^4} \sum_{i=1}^n \sum_{\substack{j=1 \\ i \neq j}}^n \sum_{\substack{l=1 \\ i \neq j \neq l}}^n \frac{1}{f^2(W_i)f^2(W_l)h^{2d}} SK^2\left(\frac{W_j - W_i}{h}\right) SK^2\left(\frac{W_j - W_l}{h}\right) \epsilon_j^4 (1 + o_p(1)) \\ &= O(n^{-1}). \end{aligned}$$

When  $i = l \neq j = m$ ,

$$V(T_{1B1}^* | Q_{(n)}) = \frac{1}{n^4} \sum_{i=1}^n \sum_{\substack{j=1 \\ i \neq j}}^n \frac{1}{f^4(W_i)h^{2d}} (SK^2\left(\frac{W_j - W_i}{h}\right))^2 \hat{\epsilon}_j^4 = O_p(n^2 h^{-d}) = o_p(n^{-1}).$$

So in all,  $V(T_{1B1}^* | Q_{(n)}) = O_p(n^{-1})$ .

Combining (i) and (ii) above, we obtain that  $T_{1B1}^* = B_1 + o_p(1)$  conditioning on  $Q_{(n)}$ .

Combining (a) and (b) above, we obtain the claim in (1).

$$\begin{aligned} (2) \quad T_3^* &= -\frac{2}{n^2} \sum_{i=1}^n \sum_{\substack{l=1 \\ l \neq i}}^n \frac{1}{f(W_i)nh^{d+1}} \sum_{j=1}^n SK\left(\frac{W_j - W_i}{h}\right) \epsilon_j^* \frac{1}{f(\omega_l; \omega_i^c)nh^{d+1}} \sum_{t=1}^n SK\left(\frac{\omega_t - \omega_l}{h}; \frac{\omega_t^c - \omega_i^c}{h}\right) \epsilon_t^* (1 + o_p(1)) \\ &= -\frac{2}{n^4} \sum_{i=1}^n \sum_{\substack{j=1 \\ i \neq j}}^n \sum_{\substack{l=1 \\ i \neq j \neq l}}^n \sum_{t=1}^n \frac{1}{f(W_i)f(\omega_l; \omega_i^c)nh^{2(d+1)}} SK\left(\frac{W_j - W_i}{h}\right) SK\left(\frac{\omega_t - \omega_l}{h}; \frac{\omega_t^c - \omega_i^c}{h}\right) \epsilon_j^* \epsilon_t^* (1 + o_p(1)) \\ &= -\frac{2}{n^4} \sum_{i=1}^n \sum_{\substack{j=1 \\ i \neq j}}^n \sum_{\substack{l=1 \\ i \neq j \neq l}}^n \sum_{\substack{t=1 \\ i \neq j, t \neq l \neq i}}^n \psi_{nijlt} \end{aligned}$$

(a) When  $i \neq j \neq l \neq t$ ,  $E(T_3^* | Q_{(n)}) = 0$  as  $E(\epsilon_j^* | Q_{(n)}) = 0$ , and

$$V(T_3^* | Q_{(n)}) = \frac{4}{n^8} \sum_{i=1}^n \sum_{\substack{j=1 \\ i \neq j}}^n \sum_{\substack{l=1 \\ i \neq j \neq l}}^n \sum_{\substack{t=1 \\ i \neq j \neq l \neq t}}^n \sum_{i'=1}^n \sum_{\substack{j'=1 \\ i' \neq j}}^n \sum_{\substack{l'=1 \\ i' \neq j \neq l}}^n \sum_{\substack{t'=1 \\ i' \neq j \neq l \neq t}}^n E(\psi_{nijlt} \psi_{ni'j'l't'} | Q_{(n)}).$$

We can show that  $E(|V(T_3^* | Q_{(n)})|) = o(n^{-2}h^{-4-d})$  with different combinations of indices in the summation, thus we apply Chebyshev's inequality to conclude that  $T_3^* = o_p(n^{-1}h^{-2-\frac{d}{2}})$ .

(b) When  $i \neq j = l \neq t$ ,

$$T_3^* = -\frac{2}{n^4} \sum_{i=1}^n \sum_{j=1}^n \sum_{t=1}^n \frac{1}{f(W_i)f(\omega_j;\omega_i^c)h^{2(d+1)}} SK\left(\frac{W_j-W_i}{h}\right) SK\left(\frac{\omega_t-\omega_j}{h}; \frac{\omega_t^c-\omega_i^c}{h}\right) \epsilon_j^* \epsilon_t^* (1 + o_p(1)).$$

Again  $E(T_3^*|Q_{(n)}) = 0$ , and we can show that  $V(T_3^*|Q_{(n)}) = O_p(n^{-4}h^{-4-d})$ , and thus  $T_3^* = o_p(n^{-1}h^{-2-\frac{d}{2}})$ .

(c) When  $i \neq j = t \neq l$ ,

$$T_3^* \equiv T_{3B}^* = -\frac{2}{n^4} \sum_{i=1}^n \sum_{j=1}^n \sum_{l=1}^n \frac{1}{f(W_i)f(\omega_l;\omega_i^c)h^{2(d+1)}} SK\left(\frac{W_j-W_i}{h}\right) SK\left(\frac{\omega_j-\omega_l}{h}; \frac{\omega_j^c-\omega_i^c}{h}\right) (\epsilon_j^*)^2 (1 + o_p(1)).$$

With tedious but similar arguments as in  $T_{1B}^*$ , we can show that

$$T_{3B}^* = -\frac{2}{nh^{2+(d-1)}} B_3(1 + o_p(1)).$$

The claim in (2) follows from (2)(a)-(2)(c) above.

(3) We perform similar arguments to obtain that

$$\begin{aligned} T_2^* &= \frac{1}{n} \sum_{i=1}^n \left[ \frac{1}{n} \sum_{j=1}^n \frac{1}{f(\omega_j;\omega_i^c)nh^{d+1}} \sum_{t=1}^n SK\left(\frac{\omega_t-\omega_j}{h}; \frac{\omega_t^c-\omega_i^c}{h}\right) \epsilon_t^* \right] \left[ \frac{1}{n} \sum_{l=1}^n \frac{1}{f(\omega_l;\omega_i^c)nh^{d+1}} \sum_{m=1}^n SK\left(\frac{\omega_m-\omega_l}{h}; \frac{\omega_m^c-\omega_i^c}{h}\right) \epsilon_m^* \right] (1 + o_p(1)) \\ &= \frac{1}{n^5} \sum_{i=1}^n \sum_{j=1}^n \sum_{t=1}^n \sum_{l=1}^n \sum_{m=1}^n \frac{1}{f(\omega_j;\omega_i^c)f(\omega_l;\omega_i^c)h^{2(d+1)}} SK\left(\frac{\omega_t-\omega_j}{h}; \frac{\omega_t^c-\omega_i^c}{h}\right) SK\left(\frac{\omega_m-\omega_l}{h}; \frac{\omega_m^c-\omega_i^c}{h}\right) \epsilon_t^* \epsilon_m^* (1 + o_p(1)) \\ &\quad \substack{i \neq j \neq t, i \neq l \neq m} \\ &= \frac{1}{n^5} \sum_{i=1}^n \sum_{j=1}^n \sum_{t=1}^n \sum_{l=1}^n \frac{1}{f(\omega_j;\omega_i^c)f(\omega_l;\omega_i^c)h^{2(d+1)}} SK\left(\frac{\omega_t-\omega_j}{h}; \frac{\omega_t^c-\omega_i^c}{h}\right) SK\left(\frac{\omega_t-\omega_l}{h}; \frac{\omega_t^c-\omega_i^c}{h}\right) (\epsilon_t^*)^2 + o_p(n^{-1}h^{-2-\frac{d}{2}}) \\ &\quad \substack{i \neq j \neq t \neq l} \\ &= T_{2B}^* + o_p(n^{-1}h^{-2-\frac{d}{2}}). \end{aligned}$$

Furthermore, we can show that  $T_{2B}^* = \frac{1}{nh^{2+d-1}} B_2(1 + o_p(1))$ . Thus, we obtain the claim in (3), which concludes the proof of Theorem 3.

□

# Bibliography

- Ait-Sahalia, Y., Bickel, P. J., Stoker, T. M., 2001. Goodness-of-fit tests for kernel regression with an application to option implied volatilities. *Journal of Econometrics* 105, 363–412.
- Azzalini, A., Bowman, A. W., 1993. On the Use of Nonparametric Regression for Checking Linear Relationships. *Journal of the Royal Statistical Society. Series B ( Methodological )* 55, 549–557.
- Chen, X., Fan, Y., 1999. Consistent hypothesis testing in semiparametric and nonparametric models for econometric time series. *Journal of Econometrics* 91, 373–401.
- de Jong, P., 1987. A central limit theorem for generalized quadratic forms. *Probability Theory and Related Fields* 75, 261–277.
- Delgado, M., Manteiga, W. G., 2001. Significance testing in nonparametric regression based on the bootstrap. *The Annals of Statistics* 29, 1469–1507.
- Dette, H., 1999. A consistent test for the functional form of a regression based on a difference of variance estimators. *Annals of Statistics* 27, 1012–1040.
- Fan, J., Gijbels, I., 1996. *Local polynomial modeling and its applications*. Chapman and Hall, London.
- Fan, J., Huang, L., 2001. Goodness-of-Fit Tests for Parametric Regression Models. *Journal of the American Statistical Association* 96, 640– 652.
- Fan, J., Huang, T., 2005. Profile likelihood inferences on semiparametric varying-coefficient partially linear models. *Bernoulli* 11 (6), 1031–1057.
- Fan, Y., Li, Q., 2002. A consistent model specification test based on the kernel sum of squares of residuals. *Econometric Reviews* 21, 337–352.
- Fan, Y., Li, Q., Min, I., 2006. A Nonparametric Bootstrap Test of Conditional Distributions. *Econometric Theory* 22, 587–613.
- Fan, Y., Li, Q., Weersink, A., 1996. Semiparametric estimation of stochastic production frontier models. *Journal of Business and Economic Statistics* 14, 460–468.
- Gozalo, P. L., Linton, O. B., 2001. Testing additivity in generalized nonparametric regression models with estimated parameters. *Journal of Econometrics* 104 (1), 1–48.
- Gu, J., Li, D., Liu, D., 2007. Bootstrap non-parametric significance test. *Journal of Nonparametric Statistics* 19, 215–230.
- Hall, P., 1984. Central Limit Theorem for Integrated Square Error of Multivariate Nonparametric Density Estimators. *Journal of Multivariate Analysis* 14, 1–16.
- Hardle, W., Mammen, E., 1993. Comparing nonparametric versus parametric regression fits. *Annals of Statistics* 21, 1926–1947.
- Hart, J. D., 1997. *Nonparametric smoothing and lack-of-fit test*. Springer, New York.

- Hjellvik, V., Yao, Q., Tjøstheim, D., 1998. Linearity testing using local polynomial approximation. *Journal of Statistical Planning and Inference* 68, 295–321.
- Hong, Y., Lee, Y.-J., 2009. A loss function approach to model specification testing and its relative efficiency to the GLR test. Manuscript, Cornell University.
- Hsiao, C., Li, Q., Racine, J., 2007. A consistent model specification test with mixed discrete and continuous data. *Journal of Econometrics* 140, 802–826.
- Korolev, I., 2018. Consistent LM Type Specification Tests for Semiparametric Models. Working paper.
- Kress, J.-P., Neumann, M. H., Yao, Q., 2008. Bootstrap tests for simple structures in nonparametric time series regression. *Statistics and its interface* 1 ((2)), 367–380.
- Lavergne, P., Vuong, Q., 2000. Nonparametric significance testing. *Econometric Theory* 16, 576–601.
- Lewbel, A., 1995. Consistent nonparametric hypothesis tests with an application to Slutsky symmetry. *Journal of Econometrics* 67 (2), 379–401.
- Li, Q., 1999. Consistent model specification tests for time series econometric models. *Journal of Econometrics* 92, 101–147.
- Li, Q., Hsiao, C., Zinn, J., 2003. Consistent specification tests for semiparametric / nonparametric models based on series estimation methods. *Journal of Econometrics* 112, 295–325.
- Li, Q., Racine, J., 2007. *Nonparametric econometrics: theory and practice*. Princeton University Press, Princeton, NJ.
- Li, Q., Wang, S., 1998. A simple consistent bootstrap test for a parametric regression function. *Journal of Econometrics* 87, 145–165.
- Linton, O. B., Nielsen, J. P., 1995. A kernel method of estimating structured nonparametric regression based on marginal integration. *Biometrika* 82, 93–100.
- Lu, Z. Q., 1996. Multivariate locally weighted polynomial fitting and partial derivative estimation. *Journal of Multivariate Analysis* 59 (2), 187–205.
- Martins-Filho, C., Yao, F., Torero, M., 2018. Nonparametric estimation of conditional value-at-risk and expected shortfall based on extreme value theory. *Econometric Theory* 34 (1).
- Masry, E., 1996. Multivariate local polynomial regression for time series: uniform strong consistency and rates. *Journal of Time Series Analysis* 17, 571–599.
- Racine, J., 1997. Consistent significance testing for nonparametric regression. *Journal of Business and Economic Statistics* 15, 369–378.
- Ruppert, D., Wand, M. P., 1994. Multivariate locally weighted least squares regression. *Annals of Statistics* 22, 1346–1370.
- Sperlich, S., Tjøstheim, D., Yang, L., 2002. Nonparametric Estimation and Testing of Interaction in Additive Models. *Econometric Theory* 18, 197–251.
- Su, L., Ullah, A., 2008. Local polynomial estimation of nonparametric simultaneous equations models. *Journal of Econometrics* 144, 193–218.
- Su, L., Ullah, A., 2012. A nonparametric goodness-of-fit-based test for conditional heteroskedasticity. Forthcoming in *Econometric Theory*.
- Ullah, A., 1985. Specification analysis of econometric models. *Journal of Quantitative Economics* 1, 187–209.
- Verbeek, M., 2008. *A guide to modern econometrics*. John Wiley & Sons.

- Wang, X., Carriere, K. C., 2011. Assessing additivity in nonparametric modelsa kernel-based method. *Canadian Journal of Statistics* 39 (4), 632–655.
- Yang, L., Park, B. U., Xue, L., Hardle, W., 2006. Estimation and testing for varying coefficients in additive models with marginal integration. *Journal of the American Statistical Association* 101, 1212–1227.
- Yao, F., Martins-Filho, C., 2015. An Asymptotic Characterization of Finite Degree U-statistics with Sample Size-Dependent Kernels: Applications to Nonparametric Estimators and Test Statistics. *Communications in Statistics - Theory and Methods* 44 (15).
- Yao, F., Ullah, A., 2013. A nonparametric R-square test for the presence of relevant variables. *Journal of Statistical Planning and Inference* 143 (9), 1527–1547.
- Yao, F., Zhang, F., Kumbhakar, S. C., 2017. Semiparametric smooth coefficient stochastic frontier model with panel data. *Journal of Business & Economic Statistics*. Forthcoming.
- Zheng, J. X., 1996. A consistent test of functional form via nonparametric estimation techniques. *Journal of Econometrics* 75, 263–289.

## Chapter 2

# Does High Leverage Ratio influence Chinese Firm Performance? A Semiparametric Stochastic Frontier Approach with Panel Data

### 2.1 Introduction

In the past decades, financial sectors have increased the productivity and efficiency of industrial firms in China by increasing their access to corporate debt (Lin et al., 2003). Excessive debt growth has been witnessed in recent years; however, the results of firms' production overcapacity and financial performance deterioration have also been observed (Maliszewski et al., 2016). Recent studies have documented that excessive total debt relative to total assets (i.e., higher debt ratio) significantly hurts firms' performance (Lang et al., 1995; Gennaioli et al., 2012; Coricelli et al., 2012).<sup>1</sup> A higher debt ratio has also been observed to negatively affect the total factor productivity (TFP) growth of Chinese industrial firms (Maliszewski et al., 2016; Jin et al., 2017). Because the change in technical efficiency is a critical part of productivity growth (Kumbhakar et al., 2000), the excessive debt ratio of Chinese firms may significantly decrease their productivity through reducing technical efficiency (i.e., the deviation of firms' output from its frontier).

The neoclassical theory by Modigliani and Miller (1958) shows that debt financing or, more broadly, external financing under perfect capital market conditions is irrelevant to firms' value. Nonetheless, the financial market in China is far from perfect due to its inability to catch up with fast-paced industry reforms and the existence of asymmetric information (Allen et al., 2005). Furthermore, firms, particularly

---

<sup>1</sup>We note that the measure of firm's performance varies in the literature. The measures commonly include ROA, ROE, profitability, TFP growth, and technical efficiency. In this paper, we extend the effect of debt to both firm's frontier (measured by net output values) and technical efficiency (measured by the mean of technical efficiency conditioning on a composite error).

non-state-owned firms, are provided limited access to debt, and incur high financial constraints relevant to their performance (Curtis, 2013; Pessarossi and Weill, 2013). A few recent studies have explored the role of debt in firms' performance measured by technical efficiency. Weill (2008) demonstrates that higher debt improves firms' efficiency only in countries with higher institutional quality. Mugeru and Nyambane (2015) observe that the productive efficiency of farms in western Australia is positively associated with short-term debt. Agostino et al. (2018) document that higher debt ratio in small and medium-sized enterprises (SMEs) significantly diminishes the positive effect of longer lending relationships on firms' productive efficiency.

Nonetheless, the role of debt in determining the production frontier has been generally overlooked in the literature. On one hand, the debt ratio may affect the frontier directly, and we call this a *neutral* effect. According to the agency cost theory, debt financing increases financial pressure on managers to perform, reducing the "free cash-flow" problem (Jensen, 1986) and thus promoting firm's output through innovation (Weill, 2008; Giannetti, 2012). This positive impact of debt may be more relevant under a sound financial market policy and institution (Levine, 1999). However, an excessive level of debt ratio likely creates the *soft budget constraint*, inducing risk-taking behavior in firm managers (i.e., a moral hazard problem) that adversely affects the firms' output and profitability (Jensen and Meckling, 1976; Garriga, 2006). Hence, the production frontier of firms may be neutrally shifted by different levels of debt ratio, and possibly in a nonlinear fashion. On the other hand, debt financing may affect production frontier indirectly through the output elasticity of physical capital and labor, which we call *non-neutral* effects. For instance, a higher level of debt ratio relaxes Chinese firms' financial constraints, permitting more resources to be allocated toward R&D investment (Lin et al., 2017). A greater R&D investment allows for updated technology in production, which can improve a firm's productivity (Zhang et al., 2012). Wakelin (2001) and Tsai and Wang (2004) demonstrate that the output elasticity of capital and labor changes at different levels of R&D. Amoroso (2015) observes that an increase in R&D investment significantly improves the output elasticity of labor. Hence, we expect that higher debt can affect the output elasticity of capital and labor in Chinese firms through R&D.

Ignoring the potential dependence of both the neutral and non-neutral effects on debt may misspecify the production frontier model, resulting in misleading technical efficiency measures and ultimately invalidating the policy implication. In this paper, we fill the gap in the literature by investigating the impact of the debt ratio of Chinese industrial firms on the production frontier and technical efficiency. The conventional linear stochastic frontier (SF) model and data envelopment analysis (DEA) have been widely applied to study China's technical efficiency in areas that include, but are not limited to, agriculture (Bhattacharyya and Parker, 1999; Chen et al., 2009; Kalirajan and Huang, 1996; Ma and Feng,



2013; Mao and Koo, 1997; Monchuk et al., 2010), banking (Chen et al., 2005; Fungáčová et al., 2013; Hsiao et al., 2015; Ariff and Luc, 2008), energy companies (Hu and Wang, 2006; Lin and Wang, 2014), hospitals (Hu et al., 2012; Yang and Zeng, 2014; Ng, 2011), the insurance industry (Yao et al., 2007), and the iron and steel industries (Wu, 1995; Movshuk, 2004; Zhou et al., 2011).

In this paper, we investigate a SF framework for three reasons. First, different from the DEA approach, the SF model allows a random error term. This feature is desirable because, in practice, a part of the output variation is likely to be unobserved due to randomness (e.g., production shocks, measurement error), and this randomness is captured by the stochastic noise in the SF model. Second, the SF model accommodates the stochastic noise as an integral part of the production technology and can further separate the noise from inefficiency. Therefore, we can obtain a meaningful measure of firm-specific inefficiency in the presence of the stochastic noise. Finally, the semiparametric specification of our SF model greatly facilitates our empirical investigation of the neutral and non-neutral effects of debt ratio.

We observe two common problems regarding modeling efficiency using Chinese data. First, many studies have implemented a two-step approach in the SF model that can lead to biased or inconsistent estimates. In the first step, generally an estimation of the efficiency score in a standard linear SF model is performed. In the second step, the efficiency is regressed on a set of explanatory variables to investigate their impact (Chen et al., 2009; Zhou et al., 2011; Wu et al., 2002). According to the assertion by Battese and Coelli (1995) and Wang and Schmidt (2002), due to potential omitted variables in the first step, the two-step approach leads to a biased technical efficiency estimate and an inconsistent estimate in the second step. The alternative is to use the one-step SF model in Battese and Coelli (1995). However, the functional form of the frontier must be specified *a priori*. The restrictions may not reveal the nonlinearity of output elasticities of inputs, resulting in model misspecification.

Second, none of the studies on China's technical efficiency using the standard SF model, according to our review of the literature, can allow for the possibility of fully efficient firms. The reason for this is that the standard SF model assumes the firm efficiency to be a random variable with a continuous density function. Hence, the probability of having zero inefficiency is zero. Kumbhakar et al. (2013) demonstrate that the failure of allowing for the possibility of full efficient firms in an SF model significantly under-estimates technical efficiency, providing a false signal to regulators regarding the improvements in efficiency.

Our motivation to allow a nonzero probability of fully efficient firms in China is inspired by the suggestive evidence from several recent studies. Zhou et al. (2011) use provincial data from 1985-2008 to document that the provinces of Heilongjiang, Hebei, and Fujian, located in the eastern region, exhibit technical efficiency scores closer to unity. Wu and Zhou (2013) and He et al. (2015) made similar

observations for non-state-owned enterprises (SOEs) in China's manufacturing firms and textile industry. However, Zhou et al. (2011) and Wu and Zhou (2013) employ a fully nonparametric regression model by Henderson and Simar (2005) with a purely two-sided random error, compared with the SF model that uses a composite error of the inefficiency and the random error. Their estimates are essentially for a regression mean, instead of the frontier. Furthermore, they measure the technical inefficiency as the difference between firms' estimated output level and the highest estimated outputs in the sample. This implicitly assumes firms with the highest estimated outputs in the sample are fully efficient, which may not be reasonable. He et al. (2015) implement the DEA approach to compute firms' efficiency score. Without the random error, this approach is the deterministic frontier in the literature. Different from the reviewed studies, we adopt the SF model, specify a flexible frontier, and allow a nonzero probability for fully efficient firms.

To model the impact of debt ratio on frontier and technical efficiency and allow for the probability of full efficiency, we adopt a newly developed smooth coefficient SF model by Yao et al. (2018a). We allow the debt to shift the frontier neutrally, and non-neutrally by altering output elasticities. We model both effects nonparametrically as smooth functions of debt to capture the nonlinear effects. According to the discussion in Zhang et al. (2012) and Zhang et al. (2018), the debt variable does not enter the frontier as a regular input (e.g., capital or labor) in the production function. In this study, we model debt as an environment variable, with its neutral effect exerting a stand-alone facilitating impact on frontier, and its non-neutral effects altering the output elasticity of regular inputs. In a conventional SF model, the debt changes only the efficiency distribution through its conditional *mean* function, that is, in the second step. We let debt influence the frontier and the efficiency through a conditional *probability* for firms to be fully efficient. Thus, we allow for a more general stochastic relationship between debt and efficiency.<sup>2</sup> We also include a linear time trend in the probability function to capture the effect of technical progress. Therefore, our model generalizes the standard SF model and the empirical version of the neoclassical production model, where firms are deemed fully inefficient in the former and fully efficient in the latter.

We employ a balanced panel of 8,501 Chinese industrial firms from 1998-2007 from the Chinese Annual Surveys of Industrial Production, maintained by the National Bureau of Statistics of China (NBSC). Using the rich firm-level data, we investigate the debt ratio's effect on production frontier and efficiency from a micro-level foundation. We further extend our dataset to cover 1998-2013, showing that the effects of debt ratio remain fairly persistent over recent years. We observe that an increase in the debt ratio significantly decreases a firm's frontier regardless of a firm's ownership, region, and industry. We show that foreign and private firms are likely to be more efficient, but SOEs and collective firms are much

---

<sup>2</sup>See Zhang et al. (2018) for more discussion on the limitation of modeling the determinants of inefficiency through only its conditional mean function.

less efficient with higher debt. As a result, regions (central and western provinces) or industries (mining and public utilities) dominated by SOEs and collectives exhibit clear efficiency loss compared with the eastern region or the manufacturing industry dominated by foreign and private firms.

From a policy perspective, our results provide insights regarding how the debt ratio should be effectively regulated to improve the production frontier and technical efficiency, particularly for SOEs facing *soft budget constraints*. On the other front, the literature (Drehmann and Juselius, 2014; Walter and Howie, 2012; Pettis, 2013) has highlighted the importance of debt ratio regulation to reduce the future banking credit risk, to prevent financial crisis and restrain macroeconomic instability. Thus, we expect that improving firms' efficiency by effectively controlling the debt ratio would contribute to properly managing the risks associated with China's economic development. He Liu, the Vice-Premier of China, emphasizes that one major task for the Chinese government in the near future is to effectively control the overall debt ratio (World Economic Forum, 2018).

The remainder of the paper is organized as follows. Section 2 details the methodology in a semiparametric SF model specification. Section 3 discusses the data and variables used in our study. Section 4 presents and discusses empirical results, and Section 5 concludes.

## 2.2 Empirical Methodology

We adopt the SF model framework popularized by Aigner et al. (1977) and Meeusen and van Den Broeck (1977). A vast body of literature has investigated modeling a production frontier and estimating technical efficiency by using SF.<sup>3</sup> The following two recent developments in the methodology are notable.

Much effort has been made to model the frontier in a flexible fashion, alleviating the risk of misspecification. Fan et al. (1996) first introduce a nonparametric frontier model, and Martins-Filho and Yao (2015) investigate its asymptotic properties and further propose a profile-likelihood based estimator for the nonparametric frontier. Kumbhakar et al. (2007) and Park et al. (2015) model all parameters in the distribution of the composite error and frontier nonparametrically. However, the fully nonparametric frontier may not be empirically feasible with many input variables due to the well-known *curse of dimensionality* problem. We are inspired by Yao et al. (2018a) and model the frontier semiparametrically as a smooth coefficient frontier. We allow the regular inputs to enter the frontier parametrically and model the impacts of the environment variable, the debt, nonparametrically.

As aforementioned, in the standard SF framework, efficient firms are inefficient because the inefficiency variable is modeled with a continuous density; thus, the probability of having zero inefficient firms is zero.

---

<sup>3</sup>See Parmeter and Kumbhakar (2014) and Kumbhakar et al. (2015) for an comprehensive review of recent development in SF models with empirical applications.

Kumbhakar et al. (2013) and Rho and Schmidt (2015) propose a “zero inefficiency stochastic frontier” (ZISF) to allow for a nonzero probability for firms to be fully efficient, which generalizes the standard SF model (all firms are inefficient) and the empirical version of the neoclassical production model (all firms are efficient). This approach is further extended in Tran and Tsionas (2015), with the probability of being fully efficient modeled as a fully nonparametric function of the environment variable. Kumbhakar et al. (2013) demonstrate through simulations and empirical applications that the standard SF method produces a significant upward bias in the technical inefficiency. Chinese firms in eastern regions or non SOEs have been shown to have high technical efficiency scores (Zhou et al., 2011; Wu and Zhou, 2013; He et al., 2015). Thus, a SF model with a nonzero probability for fully efficient firms is more appropriate in our study.

We adopt a semiparametric smooth coefficient stochastic production frontier model (SC-ZISF) by Yao et al. (2018a) as

$$Y_{it} = \alpha(D_{it}) + K_{it}\beta_K(D_{it}) + L_{it}\beta_L(D_{it}) + \epsilon_{it} \quad (2.1)$$

where  $i = 1, \dots, n$  and  $t = 1, \dots, T$  represent firms and years, respectively;  $Y_{it}$  represents the firm's output;  $K_{it}$  represents the firm's physical capital;  $L_{it}$  represents the firm's labor; and  $D_{it}$  represents the firm's debt ratio, which is our environment variable. The smooth coefficient frontier is semiparametric in nature, and this allows the regular inputs ( $K_{it}$  and  $L_{it}$ ) to enter the frontier linearly, and the environment variable ( $D_{it}$ ) affects the frontier nonlinearly through the neutral effect  $\alpha(\cdot)$  and the non-neutral effects ( $\beta_K(\cdot)$  and  $\beta_L(\cdot)$ ); all three are smooth functions of  $D$ . Because all smooth functions are univariate, the requirement of convergence speed on data is much less demanding than the full nonparametric frontier.

We further rewrite (2.1) as

$$Y_{it} = \alpha(D_{it}) + X_{it}^\top \beta(D_{it}) + \epsilon_{it} \quad (2.2)$$

where  $X_{it} = [K_{it}, L_{it}]^\top$ , and  $\beta(D_{it}) = [\beta_K(D_{it}), \beta_L(D_{it})]^\top$ . We follow Kumbhakar et al. (2013) to assume two regimes for firms to operate. Correspondingly, the composite error  $\epsilon_{it} = v_{it}$  is used for the fully efficient regime with probability  $p(Z_{it})$ , and  $\epsilon_{it} = v_{it} - u_{it}$  is for the inefficient regime with probability  $1 - p(Z_{it})$ , where the random noise is  $v_{it} \sim N(0, \sigma_v^2)$ , and the inefficiency term is  $u_{it} \sim |N(0, \sigma_u^2)|$ .  $p(Z)$  can depend on the environment variables  $Z = [D, R]^\top$ , where  $R$  is a time trend capturing the effect of technical progress. We follow Yao et al. (2018a) to use a logistic function for  $p(\cdot)$ ,

$$p(Z_{it}) = \frac{\exp(D_{it}\gamma_D + R_{it}\gamma_R)}{1 + \exp(D_{it}\gamma_D + R_{it}\gamma_R)} \quad (2.3)$$

where  $R_{it} = t$ . Notably, we adopt a common frontier for both regimes to be  $\alpha(D_{it}) + X_{it}^\top \beta(D_{it})$ . Furthermore,  $D$  and  $R$  influence the technical efficiency through altering the probability for firms to be fully efficient.

We adopt the three-step procedures in Yao et al. (2018a) to conduct the estimation. First,

$$Y_{it} - E(Y|D_{it}) = (X_{it} - E(X|D_{it}))^\top \beta(D_{it}) + \epsilon_{it} - E(\epsilon|D_{it}).$$

Because  $E(\tilde{\epsilon}_{it}|D_{it}) = 0$  for  $\tilde{\epsilon}_{it} = \epsilon_{it} - E(\epsilon|D_{it})$ , we can estimate  $\beta(D_{it})$  in (2.2) provided the unknown  $E(Y|D)$  and  $E(X_j|D)$  can be replaced with their local linear estimates  $\pi_Y(D)$  and  $\pi_{X_j}(D)$ , respectively. Thus, using  $\tilde{Y}_{it} = Y_{it} - \pi_Y(D_{it})$  as the dependent variable, and  $\tilde{X}_{j,it} = X_{j,it} - \pi_{X_j}(D_{it})$  as the independent variable, we follow step 1 (their equation (2)) in Yao et al. (2018a) to obtain a local linear smooth coefficient estimate  $\hat{\beta}(D) = (\hat{\beta}_K(D), \hat{\beta}_L(D))^\top$ . Second, because  $\epsilon_{it} = v_{it} - u_{it}(1 - I(u_{it} = 0))$ , we obtain

$$E(\epsilon_{it}|D_{it}) = -E(u_{it}I(u_{it} > 0)|D_{it}) = -(1 - p(Z_{it}))\sqrt{\frac{2\sigma_u^2}{\pi}} = -E(u_{it}|D_{it}) \equiv -\mu(D_{it}; \theta_0).$$

With the parameter vector  $\theta_0 = (\sigma_u^2, \sigma_v^2, \gamma_D, \gamma_R)$ , we construct  $\hat{\epsilon}_{it}(\theta) = \tilde{Y}_{it} - \tilde{X}_{it}^\top \hat{\beta}(D_{it}) - \mu(D_{it}; \theta)$ . The distributional assumptions on  $v$  and  $u$  allow us to write the conditional density of  $\epsilon_i$  given  $X_i = \{X_{it}\}_{t=1}^T$  and  $Z_i = [D_i, R_i]^\top$  as  $h(\epsilon_i; X_i, Z_i, \theta_0)$ . With the constructed observations, we then estimate  $\theta$  by  $\hat{\theta}$  through step 2 (their equation (3)) in Yao et al. (2018a). Finally, given  $\hat{\theta}$ , we obtain the estimate  $\mu(D_{it}; \hat{\theta})$ , and  $\tilde{Y}_{it} = Y_{it} + \mu(D_{it}; \hat{\theta})$ . Because  $Y_{it} + \mu(D_{it}; \theta_0) = (1, X_{it}^\top)\delta(D_{it}) + \tilde{\epsilon}_{it}$  and  $E(\tilde{\epsilon}_{it}|X_{it}, D_{it}) = 0$ , we follow step 3 in Yao et al. (2018a) to perform another local linear smooth coefficient estimation of  $\tilde{Y}_{it}$  on  $(1, X_{it}^\top)$  to obtain our final estimates  $\check{\delta}(D) = [\check{\alpha}(D), \check{\beta}_K(D), \check{\beta}_L(D)]^\top$ . Notably, our estimates,  $\hat{\beta}(D)$  in step one and  $\check{\delta}(D)$  in step three, depend on a level of debt ratio  $D$ , illustrating the potential heterogeneous effects of the debt ratio on the frontier.

To implement the estimators, we use a Gaussian kernel function. In the first step,  $h = \hat{\sigma}_D^2(nT)^{-\frac{1}{5}-\delta}$  is an under-smoothing *rule-of-thumb* bandwidth, and we use it in the second step to estimate  $\theta$ , where  $\delta = 0.03$ , and  $\hat{\sigma}_D^2$  is the sample standard deviation of  $D$ . According to the comments in Yao et al. (2018a), our choice of  $h$  ensures that the bias generated in the first step does not affect the  $\theta$  estimates asymptotically in the second step. In the third step, we use the bandwidth as  $h_{cvls}$  based on a cross-validation that minimizes the integrated mean squared error:

$$h_{cvls} = \underset{h}{argmin} \sum_{i=1}^n \sum_{t=1}^n [\tilde{Y}_{it} - \check{\alpha}_{-i}(D_{it}) - X_{it}^\top \check{\beta}_{-i}(D_{it})]^2 \quad (2.4)$$

where  $\check{\alpha}_{-i}(D_{it})$  and  $\check{\beta}_{-i}(D_{it}) = [\check{\beta}_{K,-i}(D_{it}), \check{\beta}_{L,-i}(D_{it})]^\top$  are *leave-one-firm-out* estimates in step three.<sup>4</sup> Because the three-step estimation is in-line with Yao et al. (2018a), we expect that the asymptotic properties for our SC-ZISF estimators would follow Theorems 1-4 in Yao et al. (2018a).

We compare our SC-ZISF model with four commonly used parametric SF models in the literatures as benchmarks, where all firms are deemed inefficient (i.e.,  $p(Z) = 0$ ). The first model is a standard Cobb-Douglas (CD) production function

$$Y_{it} = \alpha_0 + \alpha_1 D_{it} + \beta_K K_{it} + \beta_L L_{it} + v_{it} - u_{it}, \quad (2.5)$$

where the neutral function in (2.5) is linear, that is,  $\alpha(D_{it}) = \alpha_0 + \alpha_1 D_{it}$ , and the output elasticity of each input is simply estimated by a constant; thus, CD is nested in (2.1). To allow for a nonlinear quadratic neutral function, that is,  $\alpha(D_{it}) = \alpha_0 + \alpha_1 D_{it} + \alpha_2 D_{it}^2$ , we consider the extended Cobb-Douglas (ECD) production function as

$$Y_{it} = \alpha_0 + \alpha_1 D_{it} + \alpha_2 D_{it}^2 + \beta_K K_{it} + \beta_L L_{it} + v_{it} - u_{it}. \quad (2.6)$$

Clearly, the ECD in (2.6) remains a special case of SC-ZISF in (2.1). We further consider a translog (TL) model as our third benchmark model

$$\begin{aligned} Y_{it} = & \alpha_0 + \alpha_1 D_{it} + \alpha_2 D_{it}^2 + \beta_K K_{it} + \beta_L L_{it} + \beta_{DK} D_{it} K_{it} + \beta_{DL} D_{it} L_{it} \\ & + \beta_{K2} K_{it}^2 + \beta_{L2} L_{it}^2 + \beta_{KL} K_{it} L_{it} + v_{it} - u_{it}. \end{aligned} \quad (2.7)$$

It is not nested in (2.1) due to the cross-product and squared terms of  $K$  and  $L$ . The last benchmark model is a restricted translog (RTL) model:

$$Y_{it} = \alpha_0 + \alpha_1 D_{it} + \alpha_2 D_{it}^2 + \beta_K K_{it} + \beta_L L_{it} + \beta_{DK} D_{it} K_{it} + \beta_{DL} D_{it} L_{it} + v_{it} - u_{it}, \quad (2.8)$$

where the restrictions  $\beta_{K2} = \beta_{L2} = \beta_{KL} = 0$  in (2.7) are imposed. RTL is nested in (2.1) because the output elasticities of inputs are simply linear functions of  $D$ . In the benchmark models,  $u_{it} \sim |N(\mu_{it}, \sigma_u^2)|$ . For the purpose of comparison, we follow Battese and Coelli (1995) to specify  $\mu_{it} = D_{it}\eta_D + R_{it}\eta_R$ ; thus, the conditional mean of  $u$  is a linear function of the debt ratio  $D$  and the time trend  $R$ . All parameters in the aforementioned benchmark SF models are estimated by MLE, as in Battese and Coelli (1995).

<sup>4</sup>A typical cross-validation bandwidth selection in cross-section data is to leave-one-*observation*-out. In panel data, we leave-one-*firm*-out to follow the argument by Henderson and Parmeter (2015). In this manner, we ensure that a country with its observations as outliers can be entirely removed, preventing bandwidth from being too large. This approach also speeds up the optimization dramatically in the panel data.

Section 3 discusses the data.

## 2.3 Data

Our data is obtained from the Chinese Annual Surveys of Industrial Production (ASIP), officially referred as the “all state-owned and all above-scale non-state owned industrial enterprise database”. Maintained by the NBSC, the ASIP covers enterprises in 31 provinces and municipalities (i.e., Beijing, Shanghai, Tianjin, and Chongqing), accounting for more than 90% of the gross industrial output of China. Three main industries in the dataset are manufacturing, mining, and public. See Appendix 1 for a list of the 2-digit Chinese Industrial Classification Code (CICC) and industry names.

Notably, yearly reports from the ASIP are available from 1998-2013. However, the 1998-2007 sample provides relatively reliable statistical measures (Cai and Liu, 2009), and has been employed intensively in recent literature (Hsieh and Klenow, 2009; Song et al., 2011; Berkowitz et al., 2017; Brandt et al., 2017; Zou et al., 2018). The main reason to exclude the ASIP data after 2008 has been the absence of main indicators (Zou et al., 2018). Firm’s value added and intermediate inputs, for instance, are officially not reported after 2008 in the ASIP, the Statistical Yearbook, and the Census Yearbook by the NBSC. As a result, retrieving value added and intermediate inputs by matching the data from different databases is infeasible. Another reason to exclude the ASIP data after 2008 is the change in sample selection criterion: the minimum annual sales of firms in the ASIP increased from RMB 5 million from 1998-2010 to RMB 20 million from 2011-2013. See Brandt et al. (2014) for a thorough discussion on the ASIP dataset.

In our study, we provide empirical evidences on the role of debt ratio on Chinese firm’s frontier and efficiency with the more reliable ASIP dataset from 1998-2007. Notably, based on recent concerns over the rapid build-up of debt ratio, extending the samples to cover the recent time period would be desirable; however, this task would be challenging because of the data limitation. Therefore, we employ two datasets in the empirical analysis: 1) ASIP 1998-2007, a conventional choice in the literature due to the superior data quality, and 2) ASIP 1998-2013, a reasonable choice to provide recent empirical evidence.

In both datasets, we measure a firm’s output,  $Y_{it}$ , as *value added*, or a firm’s net output values. Using value added for the output has become a common choice in the literature (Wu, 1995; Movshuk, 2004; Fu et al., 2008; Zhang et al., 2012). Using the total output values, instead of value added, does not use the final product’s market demand as guidances, which may bias the evaluation of SOEs’ performance in our sample.<sup>5</sup> In dataset 1), we obtain the *value added* directly from the ASIP, computed based on the

---

<sup>5</sup>See Fu et al. (2008). In addition, using firm output as a measure of  $Y_{it}$  typically requires production inputs to have capital, labor, and raw materials. In this set-up, the multicollinearity issue appears and affects the convergence of MLE estimation in our second step and the parametric SF models, making the empirical results unreliable. A similar issue is also pointed out by Movshuk (2004).

*production approach:*

Value added (production approach) = gross output – intermediate inputs + value-added tax.

In dataset 2), by contrast, *value added* based on the production approach is unavailable after 2008 and also cannot be computed because of the unreported intermediate inputs. Thus, we follow the suggestion by Brandt et al. (2014) and perform an alternative computation of the *value added* on the basis of the *income approach*:

Value added (income approach) = labor compensation + net indirect taxes  
+ firm operating profit + capital depreciation,

where labor compensation = total wage + total labor benefit, and net indirect taxes = sales tax + managerial tax + value added tax – subsidy income. Notably, total wage and subsidy income are not reported from 2009-2010, and total benefit is not reported from 2009-2013. We thus approximate all unreported data by using the average of the previous 2-year observations.

Our environment variable is the *debt ratio*,  $D_{it}$ , defined as a firm's total debt divided by its total assets. Our production inputs include the real *capital*,  $K_{it}$ , measured by the total fixed physical capital, and *labor*,  $L_{it}$ , the number of employees.  $Y_{it}$  and  $K_{it}$  are measured in thousands of RMB.  $Y_{it}$ ,  $K_{it}$ , and  $L_{it}$  are in natural log; thus, the smooth coefficient functions  $\beta_K(\cdot)$  and  $\beta_L(\cdot)$  in (1) can be interpreted as the output elasticity of  $K$  and  $L$ , respectively, which vary with  $D$ . By following Chen and Guariglia (2013), we further compute the real variables  $Y_{it}$  and  $K_{it}$  by using the producer price index from China's 60-year statistics yearbook by the NBSC. To alleviate the impact of firm entry and exit, we use a balanced panel with incumbent firms. To mitigate the effect of outliers, we drop firms within the highest and lowest one percentile of all variables. We drop observations with obvious measurement error, such as a negative value on employees or debt ratio. Our final data comprises 8,501 incumbent firms from the ASIP 1998-2007 and 85,010 firm-year observations, and 2,205 incumbent firms from the ASIP 1998-2013 and 35,328 firm-year observations.<sup>6</sup>

We further address the potential heterogeneous effect of debt ratio using ASIP 1998-2007 due to its rich observation and superior data quality. First, the debt ratio may play a different role in frontier and technical efficiency for firms with different ownership. Recently, several studies have observed that SOEs with soft budget constraints in China have operated with low technical efficiency compared with

---

<sup>6</sup>The smaller sample size in ASIP 1998-2013 is due to the change of firm scale since 2011, when firms with annual sales at or above 20 million RMB could be included.



non-SOEs (Fu et al., 2008; Movshuk, 2004; Wu, 1995; Wu and Zhou, 2013). We follow Guariglia et al. (2011) and Chen and Guariglia (2013) to classify four ownerships based on firms' share of paid-in capital. SOEs and collective enterprises are firms with more than 50% of the firm's paid-in capital owned by the state and by urban or rural communities, respectively. Private firms have more than 50% of paid-in capital controlled by private and legal individuals.<sup>7</sup> Finally, foreign ownership is defined as firms with more than 50% of paid-in capital from Hongkong, Macau, and Taiwan, or foreign countries.

Second, the heterogeneous impact of debt ratio may also exist across different regions. We group firms into the eastern, central, and the western regions based on the Region Classification Code from the NBSC. The eastern area includes 11 provinces and accounts for 13.5% of total area of China's territory. The plentiful natural resources and coastline have resulted in the eastern provinces becoming the most developed area of the economy since the outset of the economic reforms. The central area has eight provinces, accounts for 29.3% of China's total area, and is known to lead primarily in the development of heavy industries due to its rich metal and non-metal resources. The western area includes 12 provinces, accounts for 56.4% of China's total area, and has slowest economic growth due to limited transportation and complex terrain. See Appendix 2 for a map of China with the three regions labeled.

Finally, the debt's impact may be different across industries. For example, firms in a manufacturing industry, such as the agriculture and food processing industry, typically have higher productive efficiency than firms in public utility (Wu et al., 2002). Thus, we split firms into manufacturing industry (CICC 13-43), mining industry (CICC 6-11), and public utility industry (CICC 44-46).

The descriptive statistics for the ASIP 1998-2007 are reported in *Panel A* of Table 2.1. Each variable is summarized based on the whole sample or sub-samples according to ownership types, regions, and industries, with their sample sizes in terms of  $(n, T)$  reported in *Panel B*. Foreign firms exhibit the highest mean of  $\log \text{value added}$  (3.908), followed by SOEs (3.672) and collective (3.363) and private (3.332) firms. Foreign, private, and collective firms are more labor-intensive, and the mean of their  $\log \text{labor}$  is 5.686, 5.441, and 5.33, respectively, higher than that of their  $\log \text{capital}$  at 4.515, 3.699, and 3.763. That SOEs possess the highest average of  $\log \text{labor}$  (5.93) and  $\log \text{capital}$  (5.266) is not surprising. On one hand, SOEs have been characterized as *iron-bowl* companies and pursue highly secured and excessive employment due to political pressure on their annual output target (Berkowitz et al., 2017); on the other hand, the allocation of corporate debt has long been biased toward SOEs (Pessarossi and Weill, 2013). Facing the *soft budget constraint*, SOEs have a relatively high debt ratio with an average of 0.517.

The descriptive statistics for ASIP 1998-2013 are also provided in Table 1 and are denoted by Whole (1998-2013). We observe that the  $\ln \text{value added}$ , computed based on the income approach, has its

---

<sup>7</sup>Legal persons are defined as various domestic institutions such as banks and research institutions. Because the goal of this ownership is to profit, it is treated as private ownership.

Table 2.1: Data Descriptive Summary Statistics

			Mean	S.D.	Min	Max
Panel A	<i>ln Value added (Y)</i>	Whole (1998-2013)	4.157	1.478	-1.911	12.387
		Whole	3.69	1.436	-5.711	12.112
		SOE	3.672	1.669	-2.343	11.563
		Foreign	3.908	1.386	-5.686	11.112
		Collective	3.363	1.183	-0.542	9.774
		Private	3.332	1.254	-1.954	10.096
		Eastern	3.667	1.397	-5.711	11.563
		Central	3.771	1.56	-3.815	12.112
		Western	3.714	1.519	-3.504	11.141
		Manu	3.675	1.404	-5.711	11.563
		Mining	4.195	1.709	-1.005	10.558
		Public	3.707	1.659	-2.343	10.702
	<i>Debt ratio (D)</i>	Whole (1998-2013)	0.531	0.432	0.066	1
		Whole	0.519	0.228	0	1
		SOE	0.517	0.223	0	1
		Foreign	0.431	0.227	0	1
		Collective	0.524	0.222	0	1
		Private	0.594	0.206	0	1
		Eastern	0.519	0.226	0	1
		Central	0.533	0.229	0	1
		Western	0.504	0.238	0	1
		Manu	0.522	0.228	0	1
		Mining	0.525	0.226	0	1
		Public	0.482	0.224	0	1
	<i>ln Capital (K)</i>	Whole (1998-2013)	4.335	1.689	0.047	12.24
		Whole	4.33	1.637	0.01	13.024
		SOE	5.266	1.663	0.151	12.749
		Foreign	4.515	1.496	0.026	10.853
		Collective	3.699	1.314	0.173	10.37
		Private	3.763	1.419	0.037	11.155
		Eastern	4.22	1.605	0.011	13.024
		Central	4.625	1.772	0.01	12.749
		Western	4.682	1.584	0.11	12.499
		Manufacturing	4.207	1.585	0.011	12.749
		Mining	4.895	2.008	0.01	12.127
		Public	5.597	1.509	0.594	13.024
	<i>ln Labor (L)</i>	Whole (1998-2013)	5.951	1.168	0	12.577
		Whole	5.597	1.108	0	12.577
		SOE	5.93	1.27	2.303	12.053
		Foreign	5.686	1.099	1.609	10.082
		Collective	5.441	0.922	3.091	9.901
		Private	5.33	0.932	2.303	9.898
		Eastern	5.509	1.073	0	12.053
		Central	5.876	1.191	2.708	12.577
		Western	5.853	1.123	2.303	11.366
		Manufacturing	5.568	1.095	0	12.053
		Mining	6.589	1.5	2.639	11.825
		Public	5.638	0.967	2.303	10.299
Panel B	Sample (n, T)	Whole (8501, 10)	Whole (1998-2013) (2205,16)			
		SOE (950, 10)	Foreign (2234, 10)	Collective (293, 10)	Private (1235, 10)	
		Eastern (6446, 10)	Central (958, 10)	Western (1055, 10)		
		Manufacturing (7625, 10)	Mining (188, 10)	Public (648, 10)		

*Note:* Panel A reports the summary statistics in ASIP 1998-2013 for whole sample (Whole (1998-2013)), and in ASIP 1998-2007 across the whole sample, four ownerships (SOEs, foreigners, collectives, privates), three regions (eastern, central, western), and three industries (manufacturing, mining, public utility). Variables  $Y$  is computed based on *production approach*, and  $Y$  and  $K$  are deflated by national-level producer price index from China 60-year statistics yearbook published by NBSC. Panel B reports the whole and each sub-sample with its corresponding firms numbers (n) and time period (T).

Table 2.2: The Share of Firms with Different Ownership across Regions and Industries: 1998-2007

		SOEs	Foreign	Collective	Private
Regions	Eastern	11.39%	36.86%	15.24%	36.51%
	Central	38.9%	9.29%	18.50%	33.31%
	Western	30.72%	29.78%	12.64%	26.86%
Industry	Manufacturing	10.14%	36.39%	13.74%	39.73%
	Mining	53.36%	0.37%	24.41%	21.86%
	Public Utility	87.01%	2.78%	3.02%	7.19%

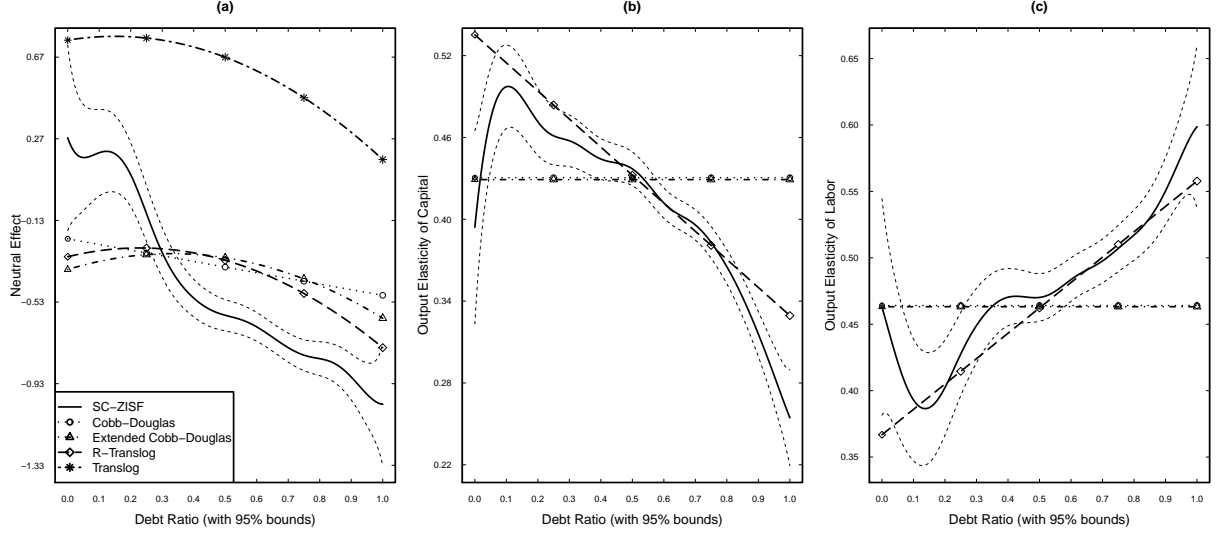
standard deviation and maximum value reasonably close to that based on the production approach (Table 1 *Panel A*). The *ln value added* from 1998-2007, obtained from two different approaches, has a correlation coefficient of 0.9534, indicating a fairly reasonable measure of value added based on the income approach. Compared with ASIP 1998-2007, this sample contains larger incumbent firms with at least 20 million RMB annual sales, with a relatively higher mean and minimum value of *ln value added*. Variables *debt ratio*, *ln capital*, and *ln labor* do not exhibit significant difference except for a relatively higher mean and standard deviation in ASIP 1998-2013.

In Table 2.2, we report the average share of firms with different ownerships across regions and industries in ASIP 1998-2007. Clearly, SOEs account for the lowest proportion of firms in the east (11.39%), the highest in the central (38.9%) and western (30.72%) where most heavy industries are located. Foreign and private firms are more dominant in eastern region, and their average share is 36.86% and 36.51%, respectively, but less dominant in the central and western areas. SOEs also largely dominate mining (53.36%) and public utility industries (87.01%). By contrast, the majority of firms in the manufacturing industry are foreign and private, with a share of 36.39% and 39.73%, respectively, whereas the share of SOEs is 10.14%. These results correspond to the observation that the highest mean of log *capital* and *labor* in Table 2.1 are observed in regions or industries where SOEs dominate.

## 2.4 Empirical Results

### 2.4.1 Whole Sample Analysis: 1998-2007

We begin our analysis on the whole sample from ASIP 1998-2007. We wonder whether a SF model with zero inefficiency should be implemented. We follow Kumbhakar et al. (2013) to test the null  $H_0: p = 0$  through a likelihood ratio (LR) test, where  $LR = -2(\ln L(\theta_{SF}) - \ln L(\theta_{ZISF}))$ , with  $L(\theta_{SF})$  ( $L(\theta_{ZISF})$ ) being the likelihood estimated from a smooth coefficient SF model without (with) zero inefficiency. Rho and Schmidt (2015) demonstrate that the LR test has the most reasonable size and power in a finite

Figure 2.1: SC-ZISF Estimation from the **Whole** Sample: 1998-2007

sample study compared with alternative test statistics. Clearly, the rejection of  $H_0$  indicates a nonzero probability of fully efficient firms through  $p(Z)$  in (2.3). Under appropriate conditions, Chen and Liang (2010) show that  $LR \xrightarrow{d} \frac{1}{2}\chi^2_{(0)} + \frac{1}{2}\chi^2_{(1)}$ , which is a 50:50 mixture  $\chi^2$  distribution. We compute the critical values by independently drawing 100,000,000 values from the mixture distribution, which provides 1.6429, 2.7066, and 5.4134 at the 10%, 5%, and 1% significant levels, respectively. Our LR test provides 51.52, clearly rejecting the SF model without full efficiency at the 1% significance level.

We first report the frontier estimation results from the SC-ZISF model on the left of *Panel A* of Table 2.3. To illustrate the impact of the debt ratio, we report  $\check{\alpha}(D)$ ,  $\check{\beta}_K(D)$ , and  $\check{\beta}_L(D)$  at the following values of  $D$  (0.1, 0.25, 0.5, 0.75, 0.9), and report the mean of neutral and non-neutral functions at the bottom of Panel A. Our results show that the frontier is neutrally shifted slightly upward only when the debt ratio is kept fairly low at approximately .15. As  $D$  reaches 0.25, the neutral function turns negative and becomes -0.096. As  $D$  gets larger, the neutral effect continues to decline. The mean of  $\check{\alpha}(D)$  (-0.4618) indicates that the neutral effect of  $D$  is negative.

To provide a vivid picture, we plot  $\check{\alpha}(D)$  (solid line) in Panel (a) of Figure 2.1 against the debt ratio, along with its 95% confident interval (short dash line, constructed with Theorem 4 in Yao et al. (2018a)). The neutral effect of  $D$  is highly nonlinear with an inverted U-shape, increasing the frontier a bit with a small value of  $D$ , and pulling down the frontier with higher level of  $D$  ( $D \geq 0.15$ ). This nonlinearity is consistent with the literature. The agency cost theory suggests that the neutral effect of debt ratio on a firm's output can be positive, because the "free cash-flow" problem (Jensen, 1986) can be reduced through the financial pressure of debt, which effectively disciplines managers' profit-maximizing behavior (Weill, 2008; Giannetti, 2012). However, the moral hazard theory argues that a higher debt ratio may easily

create a *soft budget constraint* and can induce firm managers to take risky projects, resulting in minimum return and adversely affecting the firms' output and profitability (Jensen and Meckling, 1976; Garriga, 2006). Given that the positive effect of debt ratio is more apparent under sound financial institutions and regulation (Levine, 1999), our results suggest a much more dominating role of moral hazard behavior in Chinese industrial firms, causing the output level to decrease sharply as the debt ratio rises.

We report the estimates of the non-neutral function  $\check{\beta}_K(D)$  at five levels of  $D$  in *Panel A* of Table 2.3. The mean of  $\check{\beta}_K(D)$  is 0.414, suggesting a positive output elasticity of physical capital. We further plot  $\check{\beta}_K(D)$  in Panel (b) of Figure 2.1. As the debt ratio increases before reaching 0.12, output elasticity of capital is significantly increased from 0.39 to 0.49, or a 25.6% increase. Beyond a level of 0.12 in the debt ratio, however,  $\check{\beta}_K(D)$  starts to and continues to decrease. Thus,  $\check{\beta}_K(D)$  also exhibits an inverted-U shape. The increasing part of the inverted-U-shape is likely due to R&D investment, which becomes increasingly available for firms with easier access to debt (Zhang et al., 2012; Lin et al., 2017). With the R&D investment, firms can adopt better technology in production, which is likely to promote outputs through the increasing output elasticity of capital. However, in the presence of a cheap capital price, the excessive debt in recent years has lead Chinese firms to over-accumulate capital, resulting in production overcapacity (Maliszewski et al., 2016; Berkowitz et al., 2017). Thus, we expect that the excessive use of debt diminishes its output elasticity, contributing to a declining  $\check{\beta}_K(D)$ .

Different from the first two,  $\check{\beta}_L(D)$  exhibits an upward trend in general, as shown in Panel (c) of Figure 2.1. The output elasticity of labor is on average 0.476.  $\check{\beta}_L(D)$  steadily increases from 0.39 to 0.55 as the debt ratio varies from 0.1 to 0.9 (*Panel A* of Table 2.3). Similar to the output elasticity of capital, the increasing  $\check{\beta}_L(D)$  may be attributed to the relaxation in financial constraint, which allows firms to invest more in R&D and promote human capital. The investment in knowledge thus promotes labor specialization, that is, workers gain expertise on a narrower range of tasks. Labor specialization improves firms' performance (Rosen, 1983) and may lead to non-diminishing marginal returns of labor. The empirical findings from Amoroso (2015) support the conjecture. Given the tight connection between the debt ratio and R&D investment in China, we expect a similar transmission mechanism behind the increasing pattern of  $\beta_L(D)$ .

We further compute the total partial effect of debt on frontier (TP(Y)) as  $\frac{\partial \check{\alpha}(D)}{\partial D} + K_{it} \frac{\partial \check{\beta}_K(D)}{\partial D} + L_{it} \frac{\partial \check{\beta}_L(D)}{\partial D}$ , with their estimates obtained in step 3. Hence, TP(Y) is a combination of the partial effect on the neutral function  $\frac{\partial \check{\alpha}(D)}{\partial D}$  and on the non-neutral functions weighted by inputs,  $K_{it} \frac{\partial \check{\beta}_K(D)}{\partial D} + L_{it} \frac{\partial \check{\beta}_L(D)}{\partial D}$ . The mean of TP(Y) reported in *Panel C* of Table 2.3 indicates an overall negative impact of -1.141 from the debt ratio.

To compare our SC-ZISF and conventional SF estimates, in Figure 2.1 we plot the neutral and non-

Table 2.3: SC-ZISF Estimation Results from the **Whole Sample**

	ASIP 1998-2007				ASIP 1998-2013			
	Debt ratio (D)	$\hat{\alpha}(D)$	$\hat{\beta}_K(D)$	$\hat{\beta}_L(D)$	Debt ratio (D)	$\hat{\alpha}(D)$	$\hat{\beta}_K(D)$	$\hat{\beta}_L(D)$
<i>Panel A</i>	0.1	0.2017* (0.1082)	0.4972*** (0.0156)	0.393*** (0.0234)	0.1	0.2044** (0.1015)	0.5991*** (0.0141)	0.364*** (0.0204)
	0.25	-0.096*** (0.0297)	0.4612*** (0.0109)	0.4276*** (0.0160)	0.25	0.048** (0.0683)	0.5388*** (0.0102)	0.3879*** (0.0142)
	0.5	-0.5947*** (0.0391)	0.437*** (0.0063)	0.4702*** (0.0092)	0.5	-0.5732*** (0.0448)	0.482*** (0.0069)	0.4662*** (0.0098)
	0.75	-0.7903*** (0.0386)	0.383*** (0.0060)	0.5071*** (0.0089)	0.75	-1.0452*** (0.0509)	0.3946*** (0.0077)	0.5592*** (0.0111)
	0.9	-0.8974*** (0.0573)	0.3158*** (0.0084)	0.5504*** (0.0129)	0.9	-1.3454*** (0.0832)	0.3188*** (0.0116)	0.6498*** (0.0172)
	Mean	-0.4618	0.4144	0.4755	Mean	-0.5624	0.4657	0.4892
<i>Panel B</i>	$\hat{\sigma}_u^2$	$\hat{\sigma}_v^2$	$\hat{\gamma}_D$	$\hat{\gamma}_R$	$\hat{\sigma}_u^2$	$\hat{\sigma}_v^2$	$\hat{\gamma}_D$	$\hat{\gamma}_R$
	8.4939*** (0.5070)	0.3646*** (0.0042)	4.6674*** (0.3952)	0.1355*** (0.0590)	6.5585*** (0.1623)	0.5976*** (0.0049)	5.6218*** (0.4984)	0.5095*** (0.0745)
<i>Panel C</i>	Mean of TP(Y)	-1.1413			Mean of TP (Y)	-1.2605		
	Mean of TP(ln PTE)	0.3534			Mean of TP(ln PTE)	0.3686		
	Mean of TP	-0.7879			Mean of TP	-0.8919		
	Log-likelihood	5392.1			Log-likelihood	4171.2		
	Bandwidth $h_1$ :	0.1035			Bandwidth $h_1$ :	0.1152		
	Firms :	8501			Firms :	2208		
	Time Span:	10			Time Span:	16		
	Total Obs:	85010			Total Obs:	35328		

*Note:* Dependent variable is the natural log of firm value added  $Y$ , computed based on production approach. The estimated neutral function  $\hat{\alpha}(D)$  and non-neutral functions  $\hat{\beta}_K(D)$  and  $\hat{\beta}_L(D)$  are plotted against debt ratio  $D$  at 0.1, 0.25, 0.5, 0.75, 0.9. Standard errors are reported beneath the estimated coefficients. \*\*\*, \*\*, \* stands for the statistically significant level at 1%, 5% and 10%, respectively.

Table 2.4: Parametric Models Estimation Results from the **Whole Sample**

Panel A	variables	ASIP 1998-2007					ASIP 1998-2013				
		Cobb-Douglas	Extended Cobb-Douglas	Restricted Translog	Translog		Cobb-Douglas	Extended Cobb-Douglas	Restricted Translog	Translog	
Panel A	constant	-0.2203*** (0.0175)	-0.3707*** (0.0207)	-0.3079*** (0.0408)	0.7532*** (0.0613)		-0.0922*** (0.0238)	-0.0826*** (0.0239)	0.1383*** (0.0402)	1.2001*** (0.0776)	
	$D$	-0.2768*** (0.0154)	0.4695*** (0.0567)	0.3761*** (0.0863)	0.2504*** (0.0859)		-0.1027*** (0.0101)	0.2546*** (0.0185)	0.6847*** (0.0533)	0.621*** (0.0599)	
	$D^2$	—	-0.7082*** (0.0533)	-0.8216*** (0.0526)	-0.8354*** (0.0518)		—	-0.0327*** (0.0027)	-0.0502*** (0.0028)	-0.053*** (0.003)	
	$K$	0.4305*** (0.0023)	0.4291*** (0.0024)	0.5353*** (0.0058)	0.2921*** (0.0134)		0.4963*** (0.0032)	0.493*** (0.0031)	0.5685*** (0.0051)	0.4935*** (0.0145)	
	$L$	0.4641*** (0.0036)	0.4632*** (0.0035)	0.3668*** (0.0090)	0.1921*** (0.0238)		0.4169*** (0.0049)	0.4285*** (0.0047)	0.3367*** (0.0074)	0.0101*** (0.0267)	
	$D \times K$	—	—	-0.2058*** (0.0103)	-0.1855*** (0.0105)		—	—	-0.1464*** (0.0074)	-0.1097*** (0.0074)	
	$D \times L$	—	—	0.1909*** (0.0161)	0.2001*** (0.0160)		—	—	0.1693*** (0.0116)	0.125*** (0.0118)	
	$K^2$	—	—	—	0.0555*** (0.0012)		—	—	—	0.0578*** (0.0014)	
	$L^2$	—	—	—	0.0308*** (0.0028)		—	—	—	0.0584*** (0.0028)	
	$K \times L$	—	—	—	-0.0480*** (0.0030)		—	—	—	-0.0804*** (0.003)	
Panel B	$\hat{\sigma}_u^2$	6.6946*** (0.4719)	5.7927*** (0.3487)	5.7212*** (0.3559)	7.8372*** (0.5788)		4.5532*** (0.3321***)	4.8543*** (0.3032)	5.3438*** (0.3642)	5.7639*** (0.3931)	
	$\hat{\sigma}_v^2$	0.6327*** (0.0039)	0.6330*** (0.0053)	0.6317*** (0.0046)	0.5986*** (0.0039)		0.5858*** (0.0047)	0.5781*** (0.0049)	0.5781*** (0.0043)	0.5412*** (0.0044)	
	$\hat{\gamma}_D$	-6.0662*** (0.5487)	-3.7055*** (0.2063)	-4.5598*** (0.2424)	-6.9043*** (0.4705)		-0.4242*** (0.0261)	-3.0354*** (0.0148)	-3.6237*** (0.0185)	-5.0317*** (0.0244)	
	$\hat{\gamma}_R$	-1.7806*** (0.1233)	-1.5882*** (0.0953)	-1.6467*** (0.1037)	-2.6961*** (0.1527)		-1.0676*** (0.0981)	-0.0714*** (0.0014)	-0.0707*** (0.0011)	-2.0681*** (0.0014)	
	Log-likelihood	11066.3	11057.8	11038.8	10872.2		16068.52	16022.11	16063.42	16919.7	
	Firms :	8501	8501	8501	8501		2208	2208	2208	2208	
	Time Span:	10	10	10	10		16	16	16	16	
Total Obs:		85010	85010	85010	85010		35328	35328	35328	35328	

neutral function estimates from SC-ZISF together with four benchmark parametric SF models discussed in Section 2. The MLE parameter estimates of the benchmark models are provided on the left of *Panel A* of Table 2.4. First, Figure 2.1 (a) shows that the neutral impact estimates in the CD model (dotted line with circle) is a linear downward sloping line with a slope of -0.277, clearly ignoring the nonlinearity. The ECD (dot-dash line with triangle) and RTL estimates (long dash line with diamond) capture the inverted-U-shape, but most of them are outside of the 95% confidence interval constructed for the SC-ZISF estimates, casting doubt on the validity of these two estimates. The turning point on  $D$  for the neutral effect to switch to a decreasing pattern is 0.332 (0.155) in ECD (RTL), which is higher than the 0.123 in SC-ZISF. Surprisingly, the TL estimate (short-long dash line with star) largely deviates from the other estimates, predicting a positive output level uniformly over the entire range of the debt ratio. Additionally, the TL's average output elasticity is three times higher or lower than the other estimates, making the comparison difficult and unreliable.

Panels (b) and (c) in Figure 2.1 show that CD and ECD provide similar output elasticities, with  $(\beta_K, \beta_L)$  estimates being (0.431, 0.464) and (0.429, 0.463), respectively. A fair portion of the RTL elasticity estimate is inside the 95% CI of our SC-ZISF. All parametric models, however, do not illustrate the nonlinearity of neutral and non-neutral functions. We do not plot the TL estimates for the elasticities due to its largely deviation from the rest, making graphical presentation difficult. Our SC-ZISF estimates show that the capital has a slightly lower average output elasticity than labor, qualitatively similar to what has reported in Movshuk (2004) and Zhang et al. (2012). Finally, the average of return-to-scale (RTS) in SC-ZISF is 0.890, and this result is fairly close to the estimates of CD (0.894), ECD (0.892), and RTL (0.895), but again, highly different from that of TL (1.144).

These results suggest that the benchmark models widely used in the literature cannot describe the frontier adequately (perhaps less so for TL in the output elasticities), and the potential cause of this inadequacy is model-misspecification. To test for the validity of a parametric model, we follow Yao et al. (2018a) to construct a bootstrap version of the standardized test statistic  $\hat{T}_n^*$  to test  $H_0: \delta(D) = \delta(D; \beta_0)$ ,  $\delta(D) = [\alpha(D), \beta_K(D), \beta_L(D)]^T$ , and  $\delta(D; \beta_0)$  are parametrically specified functions known up to a vector of constants  $\beta_0$ , including those from CD, ECD, RTL, and TL. We estimate the alternative with our SC-ZISF approach. Using 299 bootstrap repetitions with a *rule-of-thumb* bandwidth, the test yields empirical p-values all smaller than 0.03; hence, we reject the parametric specifications considered at a significance level of 0.05.

We plot the density of the composite error estimates from SC-ZISF (with a mode of -0.302) and three parametric models - CD, RTL and TL (with mode average of -0.44, and clearly they are close to each other) in the left panel of Figure 2.2. We do not include the ECD estimates because they are nearly



identical to the estimates of CD. Clearly, a fair proportion of firms operate with inefficiency. The left of *Panel A* in Table 2.4 shows that the estimate of  $\sigma_v^2$  across the four parametric models is fairly similar and ranges from 0.60 in the TL model to 0.63 in the other three models. The  $\hat{\sigma}_u^2$  estimate in the parametric models is much higher, with the lowest estimate being 5.72 in RTL and the highest estimate being 7.84 in TL. By contrast, SC-ZISF in Table 2.3 provides a lower  $\hat{\sigma}_v^2 = 0.37$ , but an even higher  $\hat{\sigma}_u^2 = 8.49$ . Hence, the different models suggest a large magnitude of inefficiency relative to the random noise in the composite error.

That  $\hat{\sigma}_u^2$  is relatively small in all parametric models relative to that of SC-ZISF is expected. We expect that firms with full efficiency are likely to be present in our sample. Given that all firms are perceived as inefficient in conventional SF models, the presence of fully efficient firms would drive down the overall level of inefficiency, leading to an under-estimated  $\sigma_u^2$ . In our SC-ZISF, we specify the full efficiency probability  $p(Z)$  as a logistic function, with the coefficient estimate of  $D$  and  $R$  reported in *Panel B* of Table 2.3. The debt ratio and  $R$  enters  $p(Z)$  significantly and positively, suggesting that higher debt ratio increases the probability of firms to become fully efficient, and a positive role of technical progress in improving firm efficiency.

Our  $\hat{p}(Z)$  estimates have a mean of 0.88 and range from 0.71 at the 10% percentile to 0.978% at the 90% percentile. Following Kumbhakar et al. (2013), we further estimate the posterior probability of being fully efficient as

$$\check{p}_{it} = \frac{(\hat{p}(Z_{it})/\hat{\sigma}_v)\phi(\hat{\epsilon}_{it}/\hat{\sigma}_v)}{(\hat{p}(Z_{it})/\hat{\sigma}_v)\phi(\hat{\epsilon}_{it}/\hat{\sigma}_v) + (1 - \hat{p}(Z_{it}))\frac{2}{\sigma}\phi(\hat{\epsilon}_{it}/\hat{\sigma})\Phi(-\frac{\hat{\epsilon}_{it}\hat{\lambda}}{\hat{\sigma}})} \quad (2.9)$$

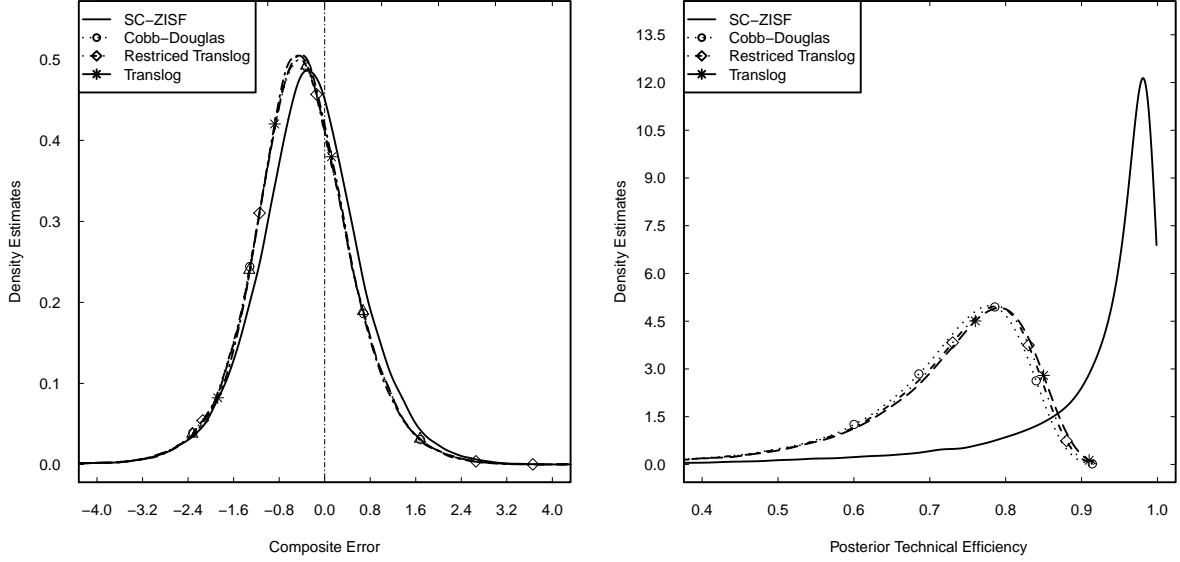
where  $\hat{\lambda} = \hat{\sigma}_u/\hat{\sigma}_v$ ,  $\hat{\sigma} = \sqrt{\hat{\sigma}_u^2 + \hat{\sigma}_v^2}$ ,  $\hat{\epsilon}_{it}$  is the estimated composite error in SC-ZISF. Clearly, debt and technical progress influence  $\check{p}_{it}$  through  $\hat{p}(Z)$ . Relative to  $\hat{p}(Z)$ , we observe that  $\check{p}_{it}$  provides a similar mean of 0.8617 but with a wider dispersion, ranging from 0.496 at the 10% percentile to 0.996 at the 90% percentile.

We further compute a posterior probability weighted technical efficiency (PTE) score following Kumbhakar et al. (2013), where  $PTE_{it} = \exp(-(1 - \check{p}_{it})\hat{u}_{it})$ , with  $\hat{u}_{it} \equiv \hat{E}(u_{it}|\epsilon_{it})$  the conditional inefficiency estimate from Jondrow et al. (1982):

$$\hat{E}(u_{it}|\epsilon_{it}) = \frac{\hat{\sigma}_u\hat{\sigma}_v}{\hat{\sigma}} \left[ \frac{\phi(\hat{\epsilon}_{it}\hat{\lambda}/\hat{\sigma})}{\Phi(-\hat{\epsilon}_{it}\hat{\lambda}/\hat{\sigma})} - (\hat{\epsilon}_{it}\hat{\lambda}/\hat{\sigma}) \right] \quad (2.10)$$

Hence, with a nonzero  $\check{p}$ , as in our case, firms in SC-ZISF should be more efficient than they would be in the conventional SF models where  $\check{p} = 0$ . We plot the density estimates of PTE for SC-ZISF and the three parametric models on the right panel of Figure 2. As expected, SC-ZISF allows the possibility of fully

Figure 2.2: Density of Estimated Composite Error and Posterior TE: 1998-2007

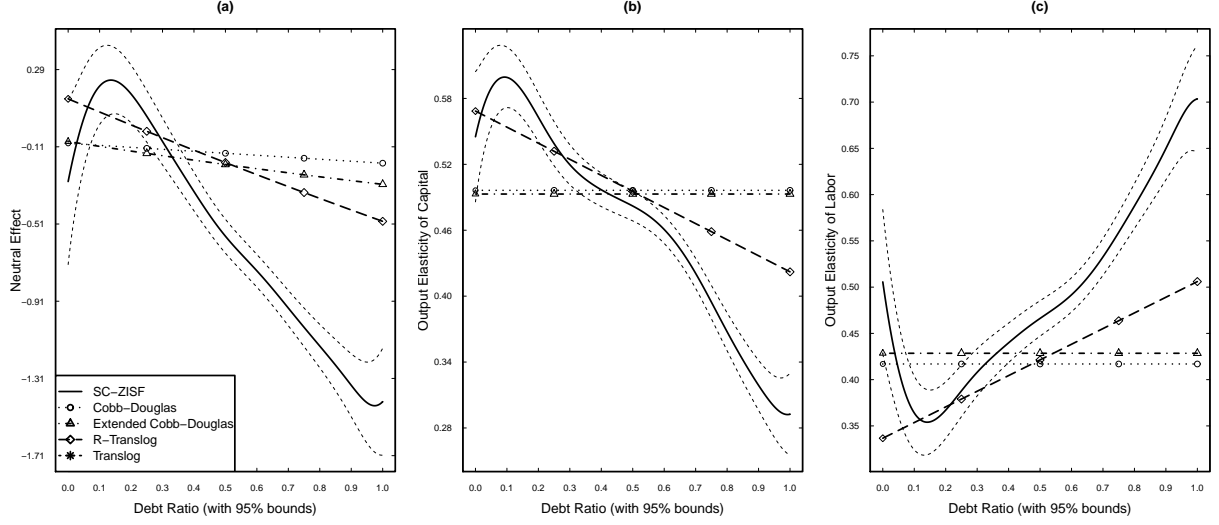


efficient firms, and the density of its PTE estimates hovers to the right of the others with a higher mean of 0.914. Moreover, SC-ZISF shows that 3,856 firms, representing approximately 45.4% of our whole sample, have PTE larger than 0.990, and 1,042 firms with PTE greater than 0.995. By contrast, the densities of PTE estimates from the parametric SF models, resembling each other in a general pattern, present an average mean of 0.715, that is, 21.8% lower than that from SC-ZISF. More important, only an average of 11 firms from parametric models have PTE greater than 0.900, and no firms has PTE greater than 0.920. Therefore, our SC-ZISF model reveals that a large portion of firms are nearly fully efficient, and clearly masked under the parametric models. This result is also in line with Kumbhakar et al. (2013) and Yao et al. (2018a), that is, parametric SF models perceive all firms as inefficient and thus largely underestimate the technical efficiency of firms.

Finally, to provide additional details on the overall effect of debt on frontier and efficiency level, we compute the total partial effect of debt on the log of PTE as  $TP(\ln PTE) = \partial \ln PTE / \partial D$ , and the total partial effect of debt on frontier and PTE as  $TP = TP(Y) + TP(\ln PTE)$ . Both measures are averaged and reported in *Panel C* of Table 2.3. Although a higher debt ratio increases firms' efficiency through PTE, it significantly reduces the production frontier, resulting in an overall output and efficiency loss.

#### 2.4.2 Whole Sample Analysis: 1998-2013

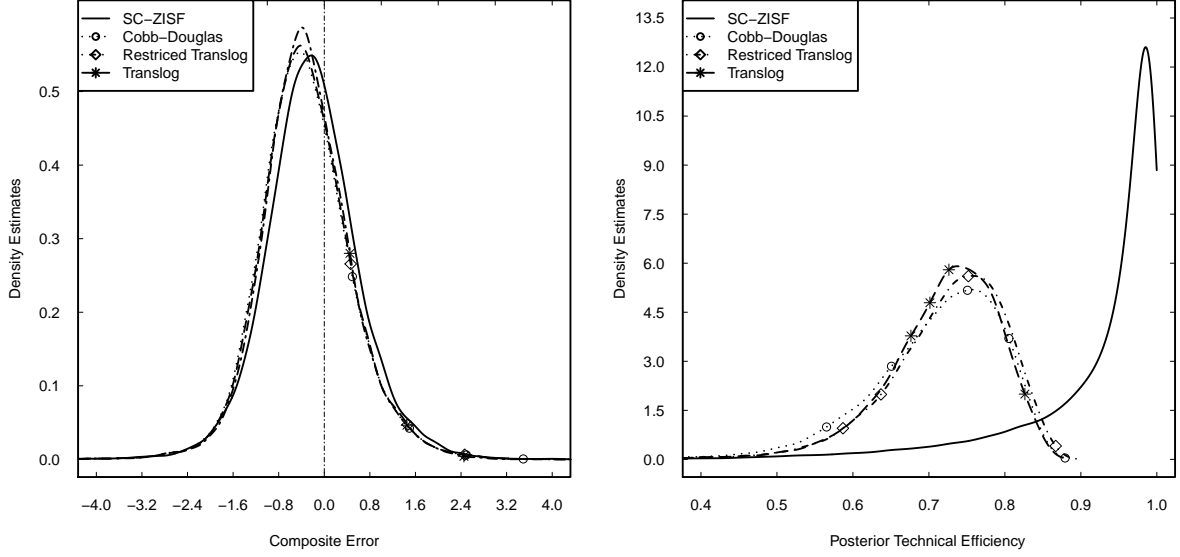
We now conduct the whole sample analysis with ASIP 1998-2013. Similar to our analysis in ASIP 1998-2007, we reject the null of a standard SF model with zero probability of full efficiency, with the LR test

Figure 2.3: SC-ZISF Estimation from the **Whole** Sample: 1998-2013

statistics being 59.28. This supports our usage of the SC-ZISF model in the sample. We report the SC-ZISF frontier estimation results in the right panel of Table 3, and plot the  $\alpha(D)$ ,  $\beta_K(D)$ , and  $\beta_L(D)$  estimates in Figure 3. We continue to observe a clear, inverted-U-shape of the neutral effect. Specifically, a firm's frontier is significantly increased from -0.29 to 0.24 when  $D$  increases to 0.13, but decreases toward -1.45 as  $D$  further increases. Relative to the neutral effect estimates in ASIP 1998-2007, we further observe that debt ratio exhibits a more negative neutral effect when  $D > 0.5$  with the ASIP 1998-2013 sample. For example, an increase in  $D$  from 0.75 to 0.90 changes the neutral effects from -1.045 to -1.345, while in the ASIP 1998-2007 sample, the corresponding effects are from -0.790 to -0.897. This result suggests that the dominating and adverse effect of debt ratio on frontier (moral hazard) is likely to persist and even intensify over time. The neutral effect of debt ratio average is decreased from -0.4618 (ASIP 1998-2007) to -0.562 (ASIP 1998-2013), a 22% drop, indicating a more detrimental neutral effect of debt ratio in recent years.

The output elasticity of  $K$  ( $L$ ) remains qualitatively similar compared to the left panel in Table 4 and Figure 1, except with a slightly higher (lower) magnitude in general. Hence, a higher debt ratio continues to decrease (increase) the return of capital (labor), and this possibly occurs through R&D. The overall effect of debt ratio, measured by  $TP(Y)$ , is -1.2605; again, this result is consistent with our previous result that a firm's frontier declines overall because of excessive debt ratio. Nevertheless, the effects of debt ratio are not well captured by the parametric models, as demonstrated in Figure 3. They generally overestimate  $\alpha(D)$  and  $\beta_K(D)$  and underestimate  $\beta_L(D)$ . The parametric SF frontier estimates are presented in the right panel in Table 4 and roughly corresponding to those in the left panel of Table 4 for ASIP 1998-2007.

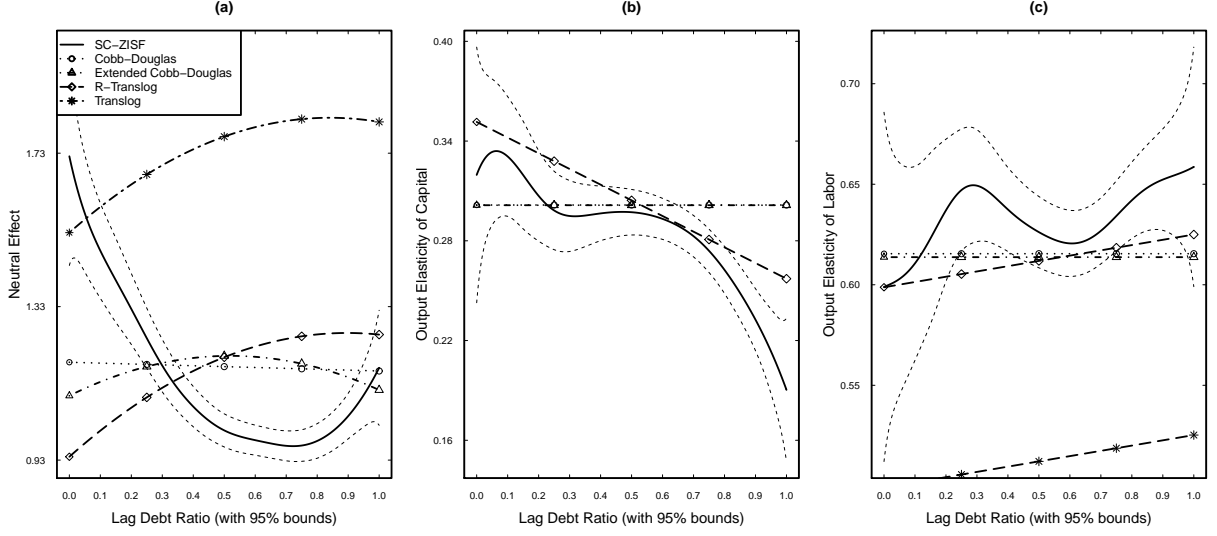
Figure 2.4: Density of Estimated Composite Error and Posterior TE: 1998-2013



The left panel in Figure 4 plots the density of estimated composite error across different SF models, which exhibits a fairly similar pattern to Figure 2 and highlights the presence of inefficient firms in the sample. This is also indicated by the larger estimate  $\hat{\sigma}_u^2$  relative to  $\hat{\sigma}_v^2$  in the right panels of Tables 3 and 4, with  $\hat{\sigma}_u^2$  being the highest in SC-ZISF, as expected. Consistent with our previous finding, the probability for firms to be fully efficient estimated in SC-ZISF, is significantly improved by higher debt ratio and technical progress. In the right panel of Figure 4, the density of the SC-ZISF PTE estimates, with a mean of 0.924, is to the right of the parametric counterparts, exhibiting a much larger mode close to one. The densities of all parametric SF PTE estimates, with a mean of 0.806 on average, still underestimate the technical efficiency. Indeed, in our SC-ZISF estimates, 1,129 firms, accounting for 51.13% of the sample, exhibit PTE larger than 0.990, and 650 firms show PTE greater than 0.995. By contrast, in the parametric estimates, no firm exhibits PTE greater than 0.900. Finally, the right side of *Panel C* in Table 3 shows that the TP is -0.892, and this result, again reflects an overall reduction in output and technical efficiency due to higher debt ratio.

In summary, our empirical findings are fairly robust and consistent across the two datasets.<sup>8</sup> Given the increasing trend of the debt ratio in recent years, we observe that the significant role of debt on frontier and technical efficiency of Chinese enterprises is likely to be persistent rather than vanishing over time. Given the relatively small sample size, concerns with the construction of the dataset in ASIP 1998-2013, and the qualitatively similar conclusion with whole sample using both datasets, we conduct

<sup>8</sup>We have also constructed dataset that covers 2008-2010, 2008-2013, and 1998-2010. The empirical results all remain qualitatively similar, and we thus omit them for brevity.

Figure 2.5: SC-ZISF Estimation with Lag of Debt Ratio from the **Whole Sample: 1998-2007**

sub-sample analyses with ASIP 1998-2007.

### 2.4.3 A Robustness Check on Whole Sample Analysis: 1998-2007

In the previous two analyses, we explore the contemporary effect of  $D_{it}$  on a firm's frontier and TE. However, the effects of debt ratio in one year may not be fully revealed until years after. To investigate the potential lag effect of  $D$  in (2.1), we employ ASIP 1998-2007 by replacing  $D_{it}$  in (2.1) with its 1-year lag value  $D_{i,t-1}$ .<sup>9</sup> We continue to reject the null of a standard SF model, with the LR test given by 43.124. Thus, we proceed by implementing our SC-ZISF model with the sample. To save space, we do not report the SC-ZISF estimates  $(\tilde{\alpha}(\cdot), \tilde{\beta}_K(\cdot), \tilde{\beta}_L(\cdot))$ , but plot them against those obtained from the four parametric models in panel (a)-(c) of Figure 2.5, respectively. Panel (a) shows that the 1-year lag debt ratio, again, neutrally shift downward the frontier except when  $D_{i,t-1} > 0.8$ . Thus, we do not find evidence on the positive neutral effect of debt ratio through its disciplining role after one year. In addition, the lag of debt ratio's neutral effect is more positive and higher in magnitude than its counterpart without lag. For instance, the average of  $\tilde{\alpha}(\cdot)$  with the lag of  $D$  is 1.143 relative to -0.462 without the lag of  $D$ . Thus, the neutral effect of debt is likely to persist and but vanish a year after. In addition, an increase in the lag of  $D$  introduces the diminishing return of  $\tilde{\beta}_K(\cdot)$  with a mean of 0.3307, which remains qualitatively similar to the estimates in panel (b) of Figure 2.1. A higher lag of  $D$  also in general improves the output elasticity of labor  $\tilde{\beta}_L(\cdot)$  with a mean of 0.6331, although the effect of which is less significant relative to panel (c) of Figure 2.3. The overall marginal impact of the lag of  $D$  is negative on the frontier, indicated by the average of  $TP(Y)$  as  $-0.827$ .

<sup>9</sup>In this case, our total observation reduces to  $(n, T) = (8, 501, 9)$

Table 2.5: Parameter Estimation Results of Lag of Debt Ratio

	SC-ZISF	Cobb-Douglas	Extended Cobb-Douglas	Restricted Translog	Translog
$\hat{\sigma}_u^2$	7.809*** (0.0397)	4.0058*** (0.7086)	4.4414*** (0.6057)	4.2533*** (0.7133)	4.9718*** (1.2007)
$\hat{\sigma}_v^2$	0.4208*** (0.0090)	0.5471*** (0.0041)	0.5444*** (0.0050)	0.5407*** (0.0038)	0.5125*** (0.0031)
$\gamma_D$	-5.3531*** (0.9006)	-3.1037*** (0.0578)	-3.2017*** (1.0120)	-3.5338*** (0.6380)	-3.9901*** (0.5069)
$\gamma_R$	0.8049*** (0.0827)	1.2378*** (0.0875)	1.2017 (0.0820)	1.2465*** (0.0741)	1.4083*** (0.0158)
Log-likelihood	7459.24	9257.71	9257.71	9250.02	9154.04
Firms :	8501	8501	8501	8501	8501
Time Span:	9	9	9	9	9
Total Obs:	76509	76509	76509	76509	76509

We report the parameter estimates of  $(\hat{\sigma}_u^2, \hat{\sigma}_v^2)$  and  $(\hat{\gamma}_D, \hat{\gamma}_R)$  in Table 2.5. Comparing to the corresponding estimates without the lag of  $D$  in *Panel B* of Table 2.3 and 2.4, we observe that debt ratio has a negative lag effect ( $\hat{\gamma}_D = -5.353$  in SC-ZISF) on firm's full efficiency probability. This is also confirmed from  $TP(\ln PTE)$ , with a mean of -0.133. The totally partial effect  $TP$  has an average of  $-0.960$ , 22% higher than the  $TP$  with  $D$ . Thus, our results suggest that the excessive debt ratio is likely to pull down the frontier in the current year, and also influence firms to be less efficient in the next year. As a result, firms have experienced a significant drop in both frontier and technical efficiency due to the lag effect of debt ratio.

#### 2.4.4 Analysis across Ownerships

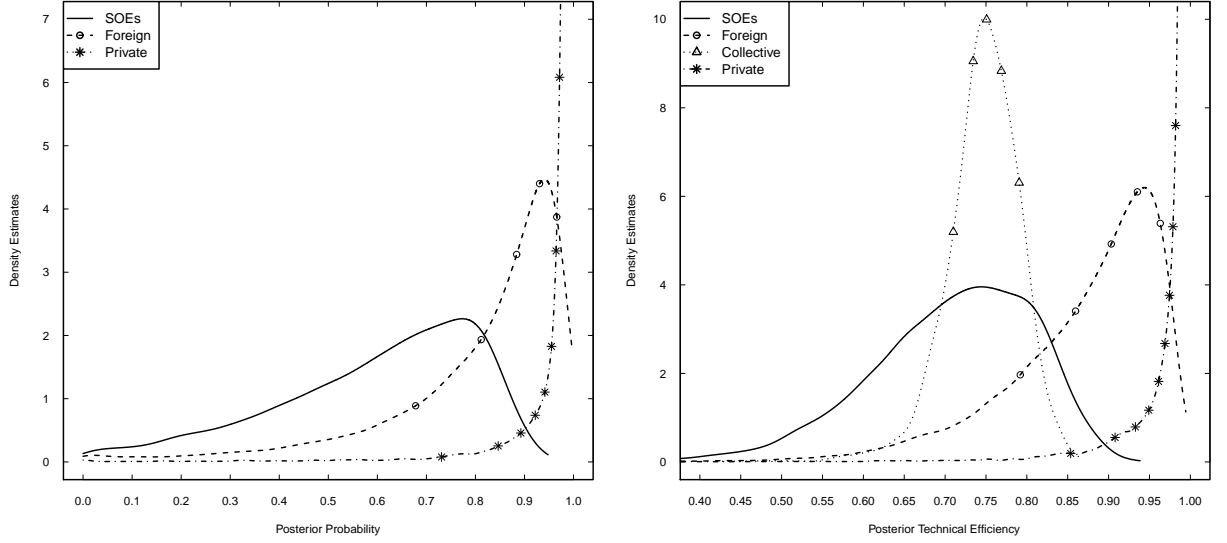
We analyze the impact of debt on frontier and efficiency across the SOEs, foreign, collective, and private firms. *Panel A* of Table 2.6 reports the estimated  $\check{\alpha}(D)$ ,  $\check{\beta}_K(D)$  and  $\check{\beta}_L(L)$  across the four ownership types. We fail to reject the null  $H_0: p = 0$  only in collective firms in *Panel C* of Table 2.6, indicating an absence of fully efficient firms in the collective firms. This result is also hinted at the insignificant SC-ZISF parameter estimates of  $\gamma_D$  (debt ratio) and  $\gamma_R$  (time trend) in *Panel B* of Table 2.6. Thus, we follow Yao et al. (2018a) to estimate a smooth coefficient SF model for the collective firms with  $p = 0$  and report its estimated frontiers and parameters  $(\hat{\sigma}_u^2, \hat{\sigma}_v^2)$  in *Panel B* of Table 2.6.

Focusing on the neutral impact  $\check{\alpha}(D)$ , the frontier of firms of different ownerships is significantly lower with a higher debt ratio, except perhaps the collective firms. We observe that SOEs' frontier neutrally decreases by a magnitude uniformly larger than the other firms. Conditioning on  $D = 0.25$ , for instance,  $\check{\alpha}(D)$  from SOEs is -1.611, whose absolute value is at least three times larger than the other firms, where the estimate for foreign, collective, and private firms is -0.220, -0.196, and -0.408, respectively. Not surprisingly, the average of the neutral effect at the bottom of *Panel A* is the largest in absolute value for SOEs (-1.584), followed by private (-0.781), foreign (-0.543), and collective (-0.052) firms. We posit

Table 2.6: Estimation Results of SC-ZISF Model based on **Ownership**

Panel A	Debt ratio (D)	SOEs			Foreign			Collective			Private		
		$\hat{\alpha}(D)$	$\hat{\beta}_K(D)$	$\hat{\beta}_L(D)$	$\hat{\alpha}(D)$	$\hat{\beta}_K(D)$	$\hat{\beta}_L(D)$	$\hat{\alpha}(D)$	$\hat{\beta}_K(D)$	$\hat{\beta}_L(D)$	$\hat{\alpha}(D)$	$\hat{\beta}_K(D)$	$\hat{\beta}_L(D)$
Panel A	0.1	-1.3379*** (0.2541)	0.5189*** (0.0415)	0.5269*** (0.0661)	-0.0045 (0.1156)	0.4769*** (0.0169)	0.356*** (0.0229)	0.0926 (0.4176)	0.3839*** (0.0596)	0.4100*** (0.0888)	-0.0362 (0.5986)	0.4545*** (0.0102)	0.3867*** (0.0320)
	0.25	-1.6106*** (0.1673)	0.5482*** (0.0307)	0.5202*** (0.0438)	-0.2201** (0.0944)	0.5044*** (0.0141)	0.378*** (0.0193)	-0.1621 (0.2792)	0.3009*** (0.0448)	0.519*** (0.0634)	-0.4083*** (0.1276)	0.4663*** (0.0035)	0.4122*** (0.0074)
	0.5	-1.6233*** (0.1089)	0.5391*** (0.0215)	0.5049*** (0.0275)	-0.5555*** (0.0774)	0.4854*** (0.0116)	0.4339*** (0.0158)	-0.2358 (0.1856)	0.3387*** (0.0311)	0.4905*** (0.0425)	-0.9483*** (0.0580)	0.4301*** (0.0017)	0.5102*** (0.0036)
	0.75	-1.6821*** (0.1452)	0.4189*** (0.0274)	0.5981*** (0.0353)	-0.7762*** (0.1023)	0.4319*** (0.0142)	0.4952*** (0.0202)	-0.2294 (0.2191)	0.295*** (0.0313)	0.4104*** (0.0450)	-1.066*** (0.0554)	0.3771*** (0.0013)	0.5459*** (0.0031)
	0.9	-1.6457*** (0.2357)	0.2844*** (0.0363)	0.6948*** (0.0484)	-1.0306*** (0.1905)	0.3895*** (0.0253)	0.559*** (0.0363)	0.7369 (0.6667)	0.2549*** (0.0454)	0.326*** (0.0703)	-1.3466*** (0.1068)	0.3436*** (0.0024)	0.6066*** (0.0060)
	mean	-1.5843	0.4675	0.5646	-0.5433	0.4581	0.448	-0.0519	0.316	0.4393	-0.7812	0.4148	0.4956
Panel B	$\hat{\sigma}_v^2$	2.6665***	(1.0020)		1.7975***	(0.1429)		0.1347***	(0.0652)		1.4230***	(0.3119)	
	$\hat{\sigma}_v$	0.7278***	(0.0109)		0.7111***	(0.0082)		0.6949***	(0.0398)		0.5681***	(0.0087)	
	$\hat{\gamma}_D$	0.6854	(1.6621)		3.5618***	(0.3205)		3.6862	(59.5857)		9.3347***	(0.9184)	
	$\hat{\gamma}_R$	3.1949***	(0.9977)		0.3801***	(0.0357)		3.9312	(84.6079)		-0.0606	(0.0642)	
Panel C	$H_0: p = 0$												
	Test	39.8198			447.256			0.371			136.135		
Panel D	p-value	0.0000			0.0000			0.271			0.0000		
	mean of TP ( $Y$ )	-0.5668			-0.2921			-0.8422			-0.5609		
	mean of TP ( $\ln PTE$ )	0.1424			0.3475			—			0.1051		
	mean of TP	-0.4243			-0.0834			-0.4641			-0.4558		
	Bandwidth $h_1$ :	0.1356			0.1306			0.1314			0.1314		
	Log likelihood	1205.4			2916.7			3810.82			1434.4		
	Firms :	950			2234			293			1235		
	Time Span:	10			10			10			10		
	Total Obs:	9500			22340			2930			12350		

Figure 2.6: Density of Posterior Probability and Posterior TE across ownerships



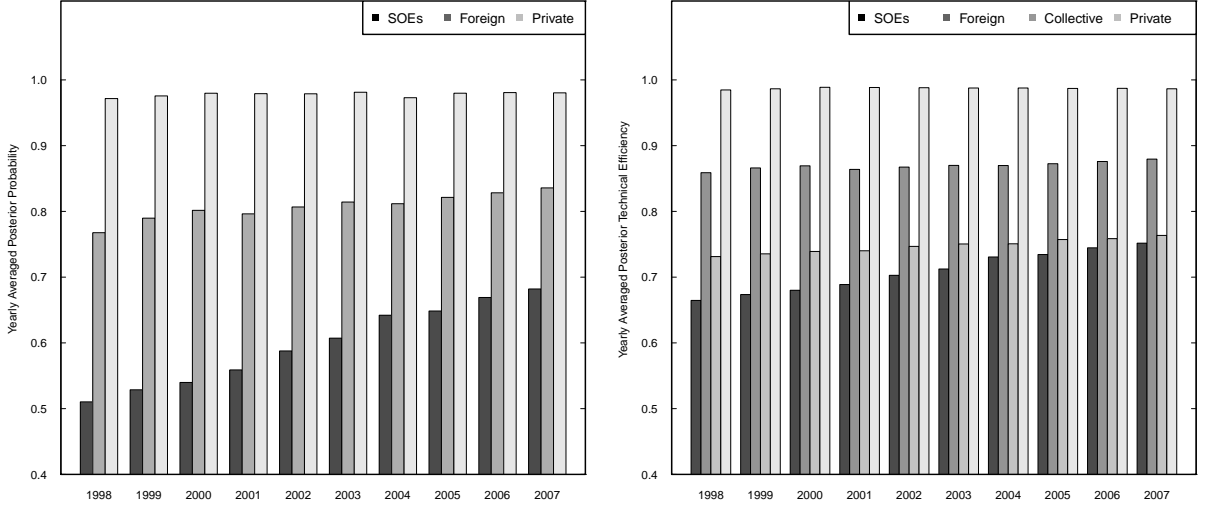
that the moral hazard issue might be present in SOEs.

We observe an invert-U-shape relationship between the debt ratio and output elasticity of capital in foreign and private firms. For example,  $\check{\beta}_K(D)$  in foreign (private) firm increases from 0.477 (0.455) to 0.504 (0.466) as  $D$  increases from 0.1 to 0.25 but steadily decreases afterward as debt increases. As in the whole sample, we posit that the result can be due to R&D's return on capital productivity (with lower debt) and the diminishing return of capital (with higher debt). We also observe the non-diminishing return of labor in SOEs, foreign, and private firms, which potentially benefit from the return of R&D on labor productivity because of the lower financial constraint. Nevertheless, the overall impact of debt on frontier is negative for all ownerships. The mean of  $TP(Y)$  is relatively small (in absolute value) in the foreign (-0.292) firms, and larger in the private (-0.561) firms, SOEs (-0.567) and collective (-0.842) firms.

From *Panel B* of Table 2.6, we observe that the  $\hat{\sigma}_v^2$  estimates are all less than one and fairly similar across different ownership types. However, SOEs have the largest  $\hat{\sigma}_u^2 = 2.667$  compared with its counterpart's estimates, 1.798, 1.192, and 1.423 for foreign, collective, and private firms, respectively. Thus, the variation of composite error in SOEs is dominated by the inefficiency term. Furthermore, debt ratio does not play a role in changing the probability of SOEs to be fully efficient. With the mean of  $TP(Y)$  and  $TP(\ln PTE)$  being -0.567 and 0.14 for SOEs, we are not surprised to observe that that excessive debt largely pulls down SOE's frontier but does not improve technical efficiency, exerting an overall negative effect on output (the mean of  $TP$  is -0.42). The technical efficiency or  $p(Z)$  in collective firms does not seem to depend on the debt ratio or technical progress. By contrast, the full efficiency probability of



Figure 2.7: Yearly Averaged Posterior Probability and PTE



foreign firms is improved with a higher debt ratio ( $\hat{\gamma}_D = 3.561$ ) and technical progress ( $\hat{\gamma}_R = 0.381$ ). Private firms' probability is mostly improved by the debt ratios ( $\hat{\gamma}_D = 9.335$ ), although not by the technical progress.

To display the difference, in Figure 2.6 we plot the density of posterior probability (left panel) and PTE (right panel) for the SOEs (solid line), foreign firms (dash line with circle), collective firms (dot line with triangle), and private firms (dot-dash line with star). We do not plot the posterior probability density for collective firms because we do not reject the null that  $p = 0$ . The  $\tilde{p}$  estimate for SOEs has a mean of 0.597, clearly exhibiting a fat tail on the left. Compared with non-SOEs, a large proportion of SOEs are far from being fully efficient. This result is in line with our previous finding that the presence of a higher debt ratio does not improve the probability for SOEs to become efficient. By contrast, the probability of private firms to be fully efficient is much higher, with a mean of 0.978, 21% higher than that of the foreign firms (0.809). The density of PTE reveals that the most inefficient firms are SOEs, with an average PTE of 0.708, lower than that of foreign (0.869) and private firms (0.987). With a mean PTE at 0.747 and a mode of 0.75, collective firms are less efficient than the other two non-SOEs. The left panel of Figure 2.7 plots the average of  $\tilde{p}$  over years across the three ownerships. Promoted by debt ratio and technical progress, private firms top the list (with an average consistently close to one from 1998 to 2007), followed by foreign firms, with its mean increasing from 0.768 in 1998 to 0.836 in 2007. As expected, the full efficiency probability for SOEs is the lowest and is mainly increased by technical progress, with its average increasing from 0.510 in 1998 to 0.682 in 2007.

The finding that SOEs have the lowest technical efficiency relative to non-SOEs is consistent with

many studies in the literature (Zheng et al., 1998; Shiu, 2002; Fu et al., 2008; He et al., 2015). The overall relatively low efficiency in SOEs may be attributed to the *soft budget constraint* problem.<sup>10</sup> The managers of SOEs typically have a strong affiliation with the government, allowing the SOEs' loans to be borrowed at a much lower rate compared with a competitive market (Ding et al., 2016). Additionally, the central and local government in China has intensively intervened in lending during economic reforms, making considerable funds available for SOEs that are not subject to close monitoring (Riedel et al., 2007). Hence, we expect that the historical *soft budget constraint* has induced the existence of a potentially severe moral hazard problem among SOEs, rather than a discipline effect. This expectation may explain that the debt shifts down the frontier neutrally by the highest magnitude but plays no role in increasing the full efficiency probability. By contrast, non-SOEs (foreign, collective, and private firms) are treated differently in the financial system and pay higher interest rates for funds than SOEs and thus incur difficult budget constraints (Guariglia et al., 2011). Hence, a higher debt may result in the associated financial pressure, which disciplines non-SOEs managers to behave as profit maximizers and helps improve the full efficiency probability.

The right panel of Figure 2.6 also reveals a wide dispersion of PTE for the SOEs. Its PTE estimates range from the minimum of 0.167 to the maximum of 0.939; thus, a small fraction of SOEs may be efficient. According to Berkowitz et al. (2017), a particular group of SOEs has tight political connections and are distinguishable from other SOEs in terms of their remarkable productivity (e.g., as high as foreign and private firms). These SOEs are known as “top central SOEs,” that is, SOEs with exceptionally high political position in comparison to “central SOEs,” which are directly supervised by the state-owned Assets Supervision and Administration Commission of the State Council (SASAC). We posit that the top central SOEs may display productivity gains through better technical efficiency. As in Berkowitz et al. (2017), we pursue a conservative measure by identifying an SOE as a top central SOE in any year if it has more than 10,000 employees and an output level exceeding RMB 1 billion. We identify 24 top central SOEs in our sample, with most of their PTE in the top 10% percentile in the PTE distribution. The PTE in the top central SOEs has a mean of 0.8818 compared with 0.7038 in the sample without the top central SOEs. Hence, the top central SOEs are highly efficient among SOEs and largely contribute to the right tail of the SOEs PTE distribution.

Nonetheless, the SOEs efficiency may have been significantly improved since 1998. Historically, the market-oriented economic policy advocated by Deng Xiaoping during his “Southern Tour Speech” induces

<sup>10</sup>Other issues have been well discussed in explaining SOEs' low production efficiency. For example, the deviation of bureaucrats' political interests (i.e., focusing on annual production target by the central government) from shareholder's interests (i.e., profit maximization) reduces the efficiency of SOEs (He et al., 2015). Additionally, the political pressure on SOEs induces SOEs managers to pursue not probability but capital accumulation with a cheap price and secured employment. As a result, the input can be seriously mis-allocated, causing the efficiency of SOEs to decrease significantly (Berkowitz et al., 2017; Hsieh and Klenow, 2009).

SOEs to dramatically expand firm size and investment. In the early of 1990s; however, a vast amount of investments were not fully capitalized to increase the output level, causing SOEs to incur huge losses and budget deficits (Fu et al., 2008). To reverse the situation, a series of loss-reduction economic reforms, known as “Grasp the Large, Release the Small,” were enacted for SOEs in 1998, resulting in a massive lay-off of surplus employees. With increasing competition against non-SOEs (mainly privates), Fu et al. (2008) document that SOEs efficiency on average has improved since 1998. We confirm this observation by plotting the yearly average of PTE estimates across different ownerships on the right panel of Figure 2.7. SOEs have steadily caught up with non-SOEs with their PTE increasing from 0.665 in 1998 to 0.752 in 2007.

Overall, we observe that a high debt ratio significantly shifts down the frontier of SOEs and all non-SOEs but improves the full efficiency probability steadily only for the foreign and private firms. Hence, an excessive use of debt financing lowers the frontier of SOEs and collective firms without enhancing their full efficiency probability, impacting the output negatively as an overall effect (i.e., TP). Economic reforms on SOEs since 1998 have effectively increased SOE’s PTE, displaying an increasing trend across years and potentially contributed by a small group of top central SOEs having notable productive efficiency.

#### 2.4.5 Analysis across Regions

Zhou et al. (2011) and Zhang et al. (2012) argue that the advanced economic development in eastern regions is due to higher production efficiency than that in the central and western regions. We illustrate the potential heterogeneous impact of the debt ratio across the eastern, central, and western regions, by providing the frontier estimates in *Panel A* of Table 2.7.

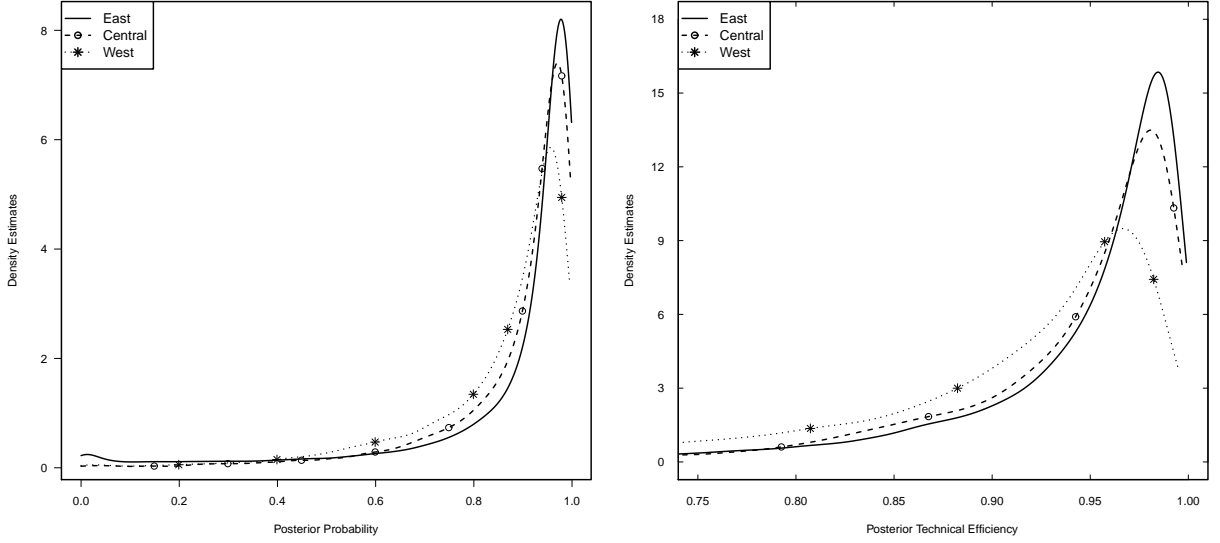
Through the neutral effect  $\alpha(D)$ , the impact of the debt ratio is negative and heterogeneous across regions. At a moderate debt ratio of 0.25, for instance,  $\check{\alpha}(D)$  is -0.386 in the eastern region, -0.673 in the central region, and -1.21 in the western region. The result may be attributed to the following: 73.37% of firms in the eastern regions are foreign and private firms, whereas 57.4% (43.36%) firms in the central (western) regions are SOEs and collective enterprises. As in Table 2.6, the neutral effect of a large debt on the frontier is more detrimental in SOEs and collective firms. We observe an inverted-U-shape for the neutral function in the eastern region, with the peak of  $\check{\alpha}(D)$  reached at a debt ratio of 0.16. However, we do not observe a similar pattern in the other two regions.

Through the non-neutral effects, the output elasticities of capital estimate,  $\check{\beta}_K(D)$ , decreases across all regions with a larger debt ratio. For example, with the debt ratio increasing from 0.1 to 0.9,  $\check{\beta}_K(D)$  decreases by 30.8%, in eastern region, and  $\check{\beta}_K(D)$  is reduced by an even larger percentage in central and western regions. The output elasticities of labor estimate,  $\check{\beta}_L(D)$ , by contrast, is significantly increased

Table 2.7: Estimation Results of the SC-ZISF Model based on Region

Panel A	Debt ratio	East			Central			West		
		$\hat{\alpha}(D)$	$\hat{\beta}_K(D)$	$\hat{\beta}_L(D)$	$\hat{\alpha}(D)$	$\hat{\beta}_K(D)$	$\hat{\beta}_L(D)$	$\hat{\alpha}(D)$	$\hat{\beta}_K(D)$	$\hat{\beta}_L(D)$
Panel A	0.1	-0.2737*** (0.0722)	0.4879*** (0.0107)	0.402*** (0.0157)	-0.3116 (0.2299)	0.4256*** (0.0298)	0.4567*** (0.0518)	-0.9841*** (0.1982)	0.5454*** (0.0272)	0.4017*** (0.0405)
	0.25	-0.3861*** (0.0519)	0.4636*** (0.0080)	0.4298*** (0.0119)	-0.6732*** (0.1625)	0.4213*** (0.0248)	0.4827*** (0.0398)	-1.2095*** (0.1488)	0.5071*** (0.0231)	0.4729*** (0.0321)
	0.5	-0.7113*** (0.0350)	0.429*** (0.0056)	0.4862*** (0.0082)	-1.092*** (0.1025)	0.3997*** (0.0178)	0.5297*** (0.0251)	-1.0645*** (0.1109)	0.4808*** (0.0169)	0.4576*** (0.0242)
	0.75	-0.8408*** (0.0396)	0.3953*** (0.0060)	0.5118*** (0.0091)	-1.2765*** (0.1201)	0.3348*** (0.0195)	0.5775*** (0.0291)	-1.1538*** (0.1276)	0.3804*** (0.0188)	0.5336*** (0.0267)
	0.9	-0.9959*** (0.0622)	0.3375*** (0.0095)	0.5644*** (0.0142)	-1.0781*** (0.1765)	0.284*** (0.0229)	0.5529*** (0.0382)	-1.171*** (0.1872)	0.3639*** (0.0273)	0.5428*** (0.0372)
	mean	-0.6655	0.4207	0.4839	-0.8878	0.3769	0.517	-1.098	0.458	0.4766
Panel B	$\hat{\sigma}_y^2$	1.5125***	(0.0543)		0.9838***	(0.1049)		1.214***	(0.1684)	
	$\hat{\sigma}_v$	0.3506***	(0.0024)		0.7542***	(0.0205)		0.858***	(0.0170)	
	$\hat{\gamma}_D$	4.7465***	(0.1295)		4.6732***	(0.6281)		4.0148***	(0.5181)	
	$\hat{\gamma}_R$	0.1476***	(0.0105)		-0.1532***	(0.0425)		0.1043***	(0.0501)	
Panel C		$H_0: p = 0$								
		Test	428.217		80.722			86.7225		
		p-value	0.0000		0.0000			0.0000		
Panel D	mean of TP ( $Y$ )	-0.5559			-0.9585			-0.3326		
	mean of TP ( $\ln PTE$ )	0.2501			0.2508			0.3037		
	mean of TP	-0.3057			-0.7078			-0.029		
		Bandwidth $h_1$ :	0.1246		0.1365			0.1373		
		Log likelihood	6327.4		1307.3			1352.6		
		Firms :	6446		958			1055		
		Time Span:	10		10			10		
		Total Obs:	64460		9580			10550		

Figure 2.8: Density of Posterior Probability and PTE across Regions



with a larger debt ratio across all regions, the largest being 40.4% in the eastern region, followed by 35.12% in the western region and 21.06% in the central region. For the eastern region, we observe that an increasing debt ratio increases  $\check{\beta}_L(D)$  by a larger percentage than the reduction in  $\check{\beta}_K(D)$ . Because a lower financial constraint significantly promotes Chinese firm R&D investment (Lin et al., 2017), we expect that firms in the eastern region with a higher debt ratio may lead to intensive R&D investment, with output elasticities of inputs improved; in this case, we observe that the improvement is focused on  $\check{\beta}_L(D)$ .<sup>11</sup> Nonetheless, the totally partial effect of the debt ratio on frontier ( $TP(Y)$ ) remains negative, with a mean of -0.556, -0.959, and -0.333 in the eastern, central, and western regions, respectively.

*Panel B* of Table 2.7 indicates that firms in the eastern region have the highest variation in inefficiency term (1.513) relative to that of the random noise (0.351), and both are significant at a 1% level. By contrast, the central region has the lowest inefficiency-to-noise ratio, with  $(\hat{\sigma}_u^2, \hat{\sigma}_v^2) = (0.9838, 0.7542)$ . The western region provides  $(\hat{\sigma}_u^2, \hat{\sigma}_v^2) = (1.214, 0.858)$ . Furthermore, we observe that a higher debt ratio significantly increases the firm's full efficiency probability across all regions. The probability also increases with time in all but the central region.

As in *Panel C*, we reject the null that  $p = 0$  for all regions, indicating that allowing for the presence of fully efficient firms is appropriate. Figure 2.8 plots the density of posterior probability (left) and PTE (right) across regions. For each region, a fairly large portion of firms possess high full efficiency probability, with the mean of the posterior probability being 0.8996 in the east (solid line), 0.8862 in the

<sup>11</sup>Zhang et al. (2012) find that the eastern region in China has a highly intensive R&D inventory compared to central and western region.

Table 2.8: Average of PTE and Posterior Probability across 31 Provinces

	East	APTE	APP	Central	APTE	APP	West	APTE	APP
1	Hainan	0.891	0.697	Jilin	0.923	0.854	Tibet	0.849	0.757
2	Guangdong	0.918	0.819	Hunan	0.928	0.863	Sichuan	0.901	0.841
3	Liaoning	0.912	0.803	Jiangxi	0.934	0.875	Inner Mongolia	0.902	0.848
4	Fujian	0.921	0.829	Henan	0.936	0.892	Guangxi	0.911	0.855
5	Tianjin	0.928	0.835	Anhui	0.937	0.887	Gansu	0.912	0.837
6	Beijing	0.933	0.852	Heilongjiang	0.942	0.887	Qinghai	0.913	0.836
7	Shanghai	0.938	0.884	Shanxi	0.943	0.899	ShaanXi	0.913	0.860
8	Shandong	0.940	0.866	Hubei	0.947	0.904	Yunnan	0.919	0.874
9	Hebei	0.941	0.858				Guizhou	0.919	0.861
10	Zhejiang	0.952	0.904				Xinjiang	0.923	0.874
11	Jiangsu	0.956	0.911				Ningxia	0.924	0.862
12							Chongqing	0.928	0.888
	mean	0.938	0.842		0.936	0.877		0.909	0.849

central (dash line with circle) and 0.8516 in the west (dot line with star). Due to the similarity of the posterior probability ( $\hat{p}$ ) density across the three regions, we implement a distribution equality test by Li (1996) to test the null hypothesis that the distribution of  $\hat{p}$  across regions is pair-wisely equal. Our test with 299 bootstrap repetitions and *rule-of-thumb* bandwidths provides test statistics  $\hat{T}_n^*$  to be 130.53 (for the east-central pair), 673.87 (for the east-west pair), and 158.92 (for the central-west pair), with all the p-values of 0.000. Thus, we conclude that the densities are distinctively different. Compared with the central and western regions, eastern firms are more likely to become fully efficient. The eastern region is also more efficient than the other two regions. The mean of PTE in the east is 0.9379, followed by 0.9366 and 0.9081 in the central and west, respectively.

In Table 2.8, we further provide the province-specific average of posterior probability (APP) and of PTE (APTE). Provinces in the east, central, and west are ranked in ascending order of their APTE score. In the eastern region, Jiangsu, Zhejiang, and Hebei are the top three provinces, with their APTE higher than 0.94, and their APP larger than 0.85. Hainan province has the lowest APTE and APP. One reason for this result might be that Hainan island has an economic development strategy that is highly focused on tourism rather than an industrial sector and financial market development. These findings are qualitatively similar to Zhou et al. (2011), that is, Zhejiang and Jiangsu provinces are technical efficient, and Hainan is less efficient relative to other eastern provinces.<sup>12</sup>

In the central region, Hubei, Shanxi, and Heilongjiang top the list, with their mean of APTE and APP higher than 0.94 and 0.88, respectively. The good performance of firms in Heilongjiang is expected. As the nation's "industrial base," most SOEs in the northeastern area (Heilongjiang and Jilin) have undergone major economic reforms since the early 1990s, with many policies enacted to attract high-tech

<sup>12</sup>We note that a direct comparison between our results and Zhou et al. (2011) is inappropriate. They measure the technical inefficiency as the difference between firms' estimated output level and the highest outputs estimated in the sample, whereas we measure the inefficiency through its density conditioning on composite error. Their frontier model is deterministic, thus does not separate inefficiency term from the random noisy. We model the probability of inefficiency explicitly in a semiparametric frontier model with the composite error structure.

industries, such as electronic manufacturing. Finally, Chongqing, Ningxia, and Xinjiang in the western region have the APTE above 0.92 and APP above 0.87. This finding is in line with Fu et al. (2008) and Zhou et al. (2011), who show that Xinjiang has many SOEs in the oil processing industry and has performed quite well since 1998. Finally, we are not surprised to observe that Tibet has the lowest APTE (0.849) and APP (.757), and this result is likely due to its complex geography and historically under-developed economy.

Overall, our results clearly indicate a significantly higher technical efficiency and full efficiency probability in the eastern region relative to the central and western regions, and this result is consistent with many studies (Zheng et al., 1998; Shiu, 2002; Fu et al., 2008; Zhou et al., 2011). We are also aware that many private and collective firms in the western and central regions were separated from SOEs in the late 1990s. Hence, we are likely to observe heterogeneous productive efficiency and full efficiency probability even within private and collective firms in each region. We suggest further exploration of this topic as a direction for further research.

#### 2.4.6 Analysis across Industries

Another notable thread in the literature demonstrates that efficiency differs across industries in China (Wu et al., 2002; Wu and Zhou, 2013). We classify industries into manufacturing, mining, and public utility based on their corresponding CICC (Appendix 1). Table 2.9 presents the estimation result from the SC-ZISF model. We first observe from *Panel B* that the probability of full efficiency in the mining industry is not a function of debt and technical progress. Furthermore, the test in *Panel C* shows that we fail to reject the null  $H_0: p = 0$  in the mining industry. Thus, we estimate a smooth coefficient SF model with  $p = 0$  as in Yao et al. (2018a), and report the frontier and parameter estimates  $(\hat{\sigma}_u^2, \hat{\sigma}_v^2)$  for the mining industry in *Panel A-B*.

We first observe that the neutral effect of debt on the frontier differs across the three industries. We observe an inverted U shape in the  $\check{\alpha}(D)$  estimates for the manufacturing industry, where the peak is reached for  $D = 0.18$ , and  $\check{\alpha}(D)$  steadily declines afterward. The inverted-U-shape is not observed from the neutral effect in the other two industries. For instance, the frontier in the mining industry is not significantly and neutrally shifted by the debt ratio. By contrast, the public utility industry's  $\check{\alpha}(D)$  generally increases with a large debt ratio.

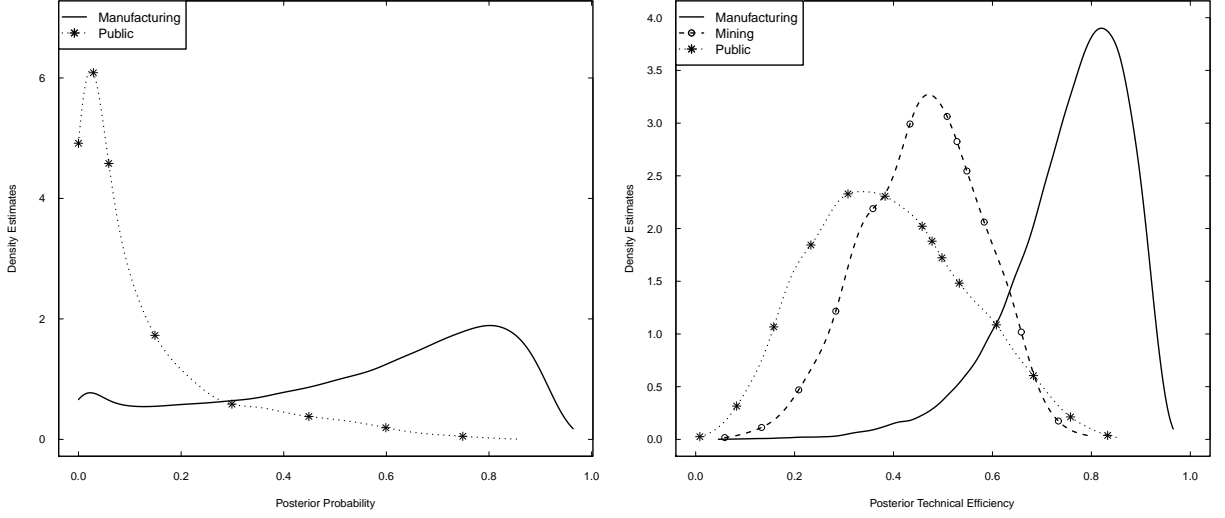
For the non-neutral effects, across all industries and for a debt ratio higher than 0.1, an increasing debt significantly reduces the output elasticity of capital  $\check{\beta}_K(D)$ , with the highest percentage reduction in the mining industry. In the manufacturing industry, we also observe a slight increase in  $\check{\beta}_K(D)$  from 0.494 to 0.505 as the debt ratio changes from 0 to 0.07. Additionally, an increase in debt ratio from

Table 2.9: Estimation Results of SC-ZISF Model based on Industry

Panel A	Debt ratio	Manufacturing			Mining			Public Utility		
		$\hat{\alpha}(D)$	$\hat{\beta}_K(D)$	$\hat{\beta}_L(D)$	$\hat{\alpha}(D)$	$\hat{\beta}_K(D)$	$\hat{\beta}_L(D)$	$\hat{\alpha}(D)$	$\hat{\beta}_K(D)$	$\hat{\beta}_L(D)$
Panel A	0.1	-0.0554 (0.2539)	0.5037*** (0.0415)	0.3657*** (0.0660)	0.9736 (0.6350)	0.4941*** (0.0634)	0.2941*** (0.1186)	-1.6567*** (0.2765)	0.841*** (0.0504)	0.283*** (0.0735)
	0.25	-0.0682 (0.1671)	0.4808*** (0.0307)	0.3983*** (0.0437)	0.2873 (0.3600)	0.4722*** (0.0513)	0.3904*** (0.0747)	-1.4884*** (0.2215)	0.8448*** (0.0394)	0.281*** (0.0572)
	0.5	-0.2552** (0.1088)	0.4475*** (0.0214)	0.4492*** (0.0275)	0.0057 (0.2822)	0.4372*** (0.0470)	0.4418*** (0.0646)	-1.5704*** (0.2116)	0.7679*** (0.0304)	0.3985*** (0.0467)
	0.75	-0.4145*** (0.1453)	0.3935*** (0.0274)	0.4969*** (0.0353)	-0.4841 (0.2705)	0.3508*** (0.0485)	0.4105*** (0.0611)	-1.4039*** (0.2716)	0.7007*** (0.0380)	0.4331*** (0.0623)
	0.9	-0.5371*** (0.2359)	0.3412*** (0.0363)	0.5379*** (0.0485)	-0.5015 (0.4774)	0.2487*** (0.0765)	0.4488*** (0.1006)	-0.6372*** (0.4112)	0.6198*** (0.0499)	0.3644*** (0.0885)
	mean	-0.276	0.4328	0.452	-0.0258	0.4018	0.4054	-1.3747	0.7511	0.3615
Panel B	$\hat{\sigma}_g^2$	1.5303***	(0.1253)		1.1154***	(0.1963)			3.1979***	(0.4070)
	$\hat{\sigma}_v^2$	0.3455***	(0.0707)		0.8679***	(0.0304)			0.7098***	(0.3019)
	$\hat{\gamma}_D$	0.5449***	(0.0615)		0.0213	(0.0615)			-4.1744***	(0.6933)
	$\hat{\gamma}_R$	0.1304***	(0.0478)		0.8427	(3.6661)			6.394***	(0.5989)
Panel C	$H_0: p = 0$									
	Test	428.217			0.3658			86.7225		
	p-value	0.0000			0.2728			0.0000		
	mean of TP ( $Y'$ )	-0.3082			-0.7827			-0.6146		
	mean of TP ( $\ln PTE$ )	0.2494			—			0.5607		
	mean of TP	-0.3323			-0.7827			-0.2346		
	Bandwidth $h_1$ :	0.1240			0.1501			0.1387		
	Log likelihood	-7502.4			-2535.15			-2557.75		
	Firms :	7625			188			648		
	Time Span:	10			10			10		
	Total Obs:	76250			1880			6480		



Figure 2.9: Density of Posterior Probability and PTE across Industries



0.1 to 0.9 significantly and monotonically increases the output elasticity of labor  $\check{\beta}_L(D)$ . Nonetheless, the positive partial effect of debt ratio on  $\beta_L(D)$  is largely offset by its negative impact on  $\alpha(D)$  and  $\beta_K(D)$ , resulting in a negative average of  $TP(Y)$  in the manufacturing (-0.326), mining (-0.783), and public utility industries (-0.615). Thus, an increase in the debt ratio pulls down the frontier across all types of industries.

From *Panel B* of Table 2.9, the public utility industry has the highest variation of inefficiency term relative to the random error, with  $(\hat{\sigma}_u^2, \hat{\sigma}_v^2) = (3.1979, 0.7098)$ .  $\hat{\sigma}_u^2$  in the public utility industry is roughly twice that for the manufacturing and mining industries. This result is partially expected because 87% of firms in the public utility industry from our sample are SOEs, whose variation of inefficiency term is higher than other ownership types (Table 2.6). The manufacturing industry has  $(\hat{\sigma}_u^2, \hat{\sigma}_v^2) = (1.5303, 0.3455)$ , and its technical efficiency and the full efficiency probability are significantly improved with a larger debt ( $\hat{\gamma}_D = 0.5449$ ) and technical progress ( $\hat{\gamma}_R = 0.1304$ ). By contrast, the efficiency of firms in the public industry only improves with technical progress. The mining industry's efficiency is not impacted by either  $D$  or  $R$ , because all firms are suggested to be inefficient. This result is consistent with the following: almost 78% of firms are either SOEs or collectives, and they are relatively inefficient (Table 2).

Based on the analysis, we expect that the technical efficiency is likely to be higher on average in the manufacturing industry than in the mining and public utility industries. We plot the density of posterior probability  $\check{p}$  on the left panel in Figure 2.9 with the mining industry omitted and PTE in the right panel. Notably, most firms of public utility are less likely to be fully efficient, with the density of posterior probability (dot line with star) far to the left, and a mean of 0.39. By contrast,  $\check{p}$  in the

manufacturing industry (solid line) ranges from 0.124 at the 10 percentile to 0.858 at 90 percentile and has the highest mean of 0.558 relative to the other industries. As expected, manufacturing firms are typically more efficient as indicated by the corresponding PTE density plot. The plot is tightly concentrated in the high efficiency range, with a mean of 0.758. Mining firms are clearly less efficient, with their PTE density (dash line with circle) largely concentrated around the mean of PTE at 0.503. We observe that the lower 10% of PTE firms comprise 73% mining companies located in the central and west, and this result is fairly consistent with the findings in Hu and Wang (2006). Public utility firms are least inefficient because their PTE density hovers in the low efficiency range with a mean of only 0.39. As indicated by the negative  $TP(Y)$ , a larger debt ratio continues to have an overall detrimental impact on the frontier of firms across all industries.

Our findings that the manufacturing industry tends to be more efficient are consistent with some early results. Wu et al. (2002) show that the manufacturing industries of agriculture and food processing (CICC 13), foodstuff (CICC 14), and transport and communication facilities (CICC 37) were the best performers in 1995 in terms of technical efficiency. By contrast, electricity firms as a part of public utility industry were the worst.

## 2.5 Conclusion

The recent growth of debt ratio increases the concerns of its impact on Chinese firm's production overcapacity and productive efficiency. However, whether and how the firm's debt ratio affects the production frontier and technical efficiency has received little attention in the literature.

In this paper, we address this topic by employing a panel of 8,501 Chinese industrial firms from 1998-2007. We further employ a panel of 2,206 firms from 1998-2013 as our second dataset to provide more recent empirical evidence for the persistent effects of debt over years. We consider the impact of debt through a semiparametric smooth coefficient SF model by Yao et al. (2018a). Different from conventional parametric SF models, our model allows the debt to affect the frontier through the neutral and non-neutral effects. We further model the firm's full efficiency probability as a function of the debt, and allow debt to affect firm's technical efficiency through the probability. Thus, our model generalizes the standard SF models and empirical version of neoclassical production models. Our model is also more robust to potential mis-specification, allowing us to investigate the impacts of debt on Chinese firms' production frontier and efficiency.

Our empirical results remain fairly robust from 1998-2007 and 1998-2013, showing that excessive debt ratio significantly pulls down the production frontier neutrally and non-neutrally through decreasing the

output elasticity of capital. Additionally, the nonlinear effects are not well captured by the restrictive model specification in standard SF models. We observe that only firms in the eastern region and in manufacturing (dominated by foreign and private firms) have the frontier moderately improved neutrally and non-neutrally when the debt ratio is maintained at a relatively low level (below 12%). However, a larger debt ratio overall declines in the frontier, and this relationship is consistently displayed across different ownerships, regions, and industries. We observe that the probability of full efficiency for foreign and private firms is strengthened by the use of debt. By contrast, SOEs and collectives are not likely to be fully efficient with higher debt. Hence, an excessive debt decreases both the frontier and efficiency of SOEs and collectives. Furthermore, we observe that the eastern region has a much higher technical efficiency relative to the central and western regions. The manufacturing industry exhibits a higher efficiency level than the mining and public utility industries where SOEs and collectives dominate. These results are expected because the eastern region and manufacturing industry are dominated by foreign and private firms.

Based on our results, we call for the Chinese government to implement policies that effectively regulate debt ratio and, in particular, minimize the moral hazard for the sake of increasing firms' frontiers and technical efficiency. Among others, we observe that the *soft budget constraint* is present in SOEs. The excessive debt decreases SOEs' production frontier by a large magnitude without improving technical efficiency. Since the loss-reduction reforms for SOEs in 1998, we observe that SOEs improve over time in terms of the average efficiency but continue to be much less efficient relative to foreign and private firms. Given the dominance of SOEs in many provinces and industries, we expect that policies that promote the effective use of debt in SOEs may help to shrink the efficiency gap across regions and industries in China.

# Bibliography

- Agostino, M., Ruberto, S., Trivieri, F., 2018. Lasting lending relationships and technical efficiency: evidence on European SMEs. *Journal of Productivity Analysis*, 1–16.
- Aigner, D., Lovell, C. A.K., Schmidt, P., 1977. Formulation and estimation of stochastic frontiers production function models. *Journal of Econometrics* 6 (1), 21–37.
- Allen, F., Qian, J., Qian, M., 2005. Law, finance, and economic growth in China. *Journal of Financial Economics* 77 (1), 57–116.
- Amoroso, S., 2015. Profits, R&D and labour: breaking the law of diminishing returns to labour. Working papers on corporate R&D and innovation, Joint Research Center.
- Ariff, M., Luc, C., 2008. Cost and profit efficiency of Chinese banks: a non-parametric analysis. *China Economic Review* 19 (2), 260–273.
- Battese, G.E., Coelli, T.J., 1995. A model for technical inefficiency effects in a stochastic frontier production function for panel data. *Empirical Economics* 20 (2), 325–332.
- Berkowitz, D., Ma, H., Nishioka, S., 2017. Recasting the iron rice bowl: the reform of China’s state-owned enterprises. *Review of Economics and Statistics* 99 (4), 735–747.
- Bhattacharyya, A., Parker, E., 1999. Labor productivity and migration in Chinese agriculture a stochastic frontier approach. *China Economic Review* 10 (1), 59–74.
- Brandt, L., Loren, Van Biesebroeck, J., Zhang, Y., 2014. Challenges of working with the Chinese NBS firm-level data. *China Economic Review* 30, 339–352.
- Brandt, L., Loren, Van Biesebroeck, J., Wang, L., Zhang, Yifan, 2017. WTO accession and performance of Chinese manufacturing firms. *American Economic Review* 107(9), 2784–2820.
- Cai, H., Liu, Q., 2009. Competition and corporate tax avoidance: evidence from Chinese industrial firms. *The Economic Journal* 119 (537), 764–795.
- Cai, Z., Fan, J., Li, R., 2000. Efficient estimation and inferences for varying-coefficient models. *Journal of the American Statistical Association* 95 (451), 888–902.
- Chen, M., Guariglia, A., 2013. Internal financial constraints and firm productivity in China: do liquidity and export behavior make a difference? *Journal of Comparative Economics* 41 (4), 1123–1140.
- Chen, X.G., Michael, S., Kym, B., 2005. Banking efficiency in China: Application of DEA to pre-and post-deregulation eras: 1993–2000. *China Economic Review* 16 (3), 229–245.
- Chen, Y., Liang, K.-Y., 2010. On the asymptotic behaviour of the pseudolikelihood ratio test statistic with boundary problems. *Biometrika* 97 (3), 603–620.
- Chen, Z., Huffman, W.E., Rozelle, S., 2009. Farm technology and technical efficiency: evidence from four regions in China. *China Economic Review* 20 (2), 153–161.

- Coricelli, F., Driffield, N., Pal, S., Roland, I., 2012. When does leverage hurt productivity growth? A firm-level analysis. *Journal of International Money and Finance* 31 (6), 1674–1694.
- Curtis, C.C., 2013. Economic reforms and the evolution of Chinas TFP in the state and the private sectors. Working paper, University of Richmond.
- Ding, S., Knight, J., Zhang, X., 2016. Does china overinvest? Evidence from a panel of Chinese firms. *The European Journal of Finance*, 1–23.
- Drehmann, M., Juselius, M., 2014. Evaluating early warning indicators of banking crises: Satisfying policy requirements. *International Journal of Forecasting* 30 (3), 759–780.
- Fan, Y., Li, Q., Weersink, A., 1996. Semiparametric estimation of stochastic production frontier models. *Journal of Business and Economic Statistics* 14, 460–468.
- Fu, F.C., Vijverberg, C.P.C., Chen, Y.S., 2008. Productivity and efficiency of state-owned enterprises in China. *Journal of Productivity Analysis* 29 (3), 249–259.
- Fungáčová, Z., Pessarossi, P., Weill, L., 2013. Is bank competition detrimental to efficiency? evidence from China. *China Economic Review* 27, 121–134.
- Garriga, J.M., 2006. The effect of relationship lending on firm performance. University Autònoma de Barcelona. Department d'Economia de l'Empresa
- Gennaioli, N., Shleifer, A., Vishny, R., 2012. Neglected risks, financial innovation, and financial fragility. *Journal of Financial Economics* 104 (3), 452–468.
- Giannetti, C., 2012. Relationship lending and firm innovativeness. *Journal of Empirical Finance* 19 (5), 762–781.
- Guariglia, A., Liu, X., Song, L., 2011. Internal finance and growth: microeconomic evidence on Chinese firms. *Journal of Development Economics* 96 (1), 79–94.
- He, Y., Chiu, Y.H., Zhang, B., 2015. The impact of corporate governance on state-owned and non-state-owned firms efficiency in China. *The North American Journal of Economics and Finance* 33, 252–277.
- Henderson, D.J., Parmeter, C.F., 2015. *Applied nonparametric econometrics*. Cambridge University Press.
- Henderson, D.J., Simar, L., 2005. A fully nonparametric stochastic frontier model for panel data. Unpublished manuscript. Retrieved from <https://sites.uclouvain.be/ISBA-Archives/ISBApublications/archive/dp2005/dp0525.pdf>
- Hsiao, C., Yan, S., Wenlong, B., 2015. Evaluating the effectiveness of China's financial reformthe efficiency of China's domestic banks. *China Economic Review* 35, 70–82.
- Hsieh, C.T., Klenow, P.J., 2009. Misallocation and manufacturing TFP in China and India. *The Quarterly Journal of Economics* 124 (4), 1403–1448.
- Hu, H.H., Qi, Q., Yang, C.H., 2012. Analysis of hospital technical efficiency in China: effect of health insurance reform. *China Economic Review* 23 (4), 865–877.
- Hu, J.L., Wang, S.C., 2006. Total-factor energy efficiency of regions in China. *Energy Policy* 34 (17), 3206–3217.
- Jensen, M., Meckling, W., 1976. Theory of the firm: managerial behavior, agency costs and ownership structure. *Journal of Financial Economics* 3 (4), 305–360.
- Jensen, M.C., 1986. Agency costs of free cash flow, corporate finance, and takeovers. *The American Economic Review* 76 (2), 323–329.
- Jin, M., Zhao, S., Kumbhakar, S. C., 2017. Financial constraints and firm productivity: evidence from Chinese manufacturing. Working paper.

- Jondrow, J., Lovell, C. A.K., Materov, I.S., Schmidt, P., 1982. On the estimation of technical inefficiency in the stochastic frontier production function model. *Journal of Econometrics* 19, 233–238.
- Kalirajan, K., Huang, Y., 1996. An alternative method of measuring economic efficiency: the case of grain production in China. *China Economic Review* 7 (2), 193–203.
- Kumbhakar, S.C., Denny, M., Fuss, M., 2000. Estimation and decomposition of productivity change when production is not efficient: a panel data approach. *Econometric Reviews* 19 (4), 312–320.
- Kumbhakar, S.C., Park, B.U., Simar, L., Tsionas, E., 2007. Nonparametric stochastic frontiers: a local maximum likelihood approach. *Journal of Econometrics* 137, 1–27.
- Kumbhakar, S.C., Parmeter, C.F., Tsionas, E.G., 2013. A zero inefficiency stochastic frontier model. *Journal of Econometrics* 172 (1), 66–76.
- Kumbhakar, S.C., Wang, H.J., Horncastle, A.P., 2015. *A Practitioner's Guide to Stochastic Frontier Analysis Using Stata*. Cambridge University Press.
- Lang, L., Ofek, E., Stulz, R.M., 1995. Leverage, investment, and firm growth. *Journal of Financial Economics*, 40 (1), 3–29.
- Levine, R., 1999. Financial development and economic growth: views and agenda. The World Bank.
- Li, Q., 1996. Nonparametric testing of closeness between two unknown distribution functions. *Econometric Reviews* 15, 261–274.
- Lin, B., Wang, X., 2014. Exploring energy efficiency in China's iron and steel industry: a stochastic frontier approach. *Energy Policy* 72, 87–96.
- Lin, J.Y., Cai, F., Li, Z., 2003. *The China miracle: development strategy and economic reform*. Chinese University Press.
- Lin, Z.J., Liu, S., Sun, F., 2017. The impact of financing constraints and agency costs on corporate R&D investment: evidence from China. *International Review of Finance* 17 (1), 3–42.
- Walter, C., Howie, F., 2012. *Red capitalism: The fragile financial foundation of China's extraordinary rise*. Singapore: John Wiley.
- Liu H., 2018. Three critical battles China is preparing to fight. Speech in World Economic Forum Annual Meeting, Davos, 2018.
- Lipton, D., 2017. Rebalancing China: International lessons in corporate debt. *China Economic Review* 10.
- Ma, S., Feng, H., 2013. Will the decline of efficiency in China's agriculture come to an end? an analysis based on opening and convergence. *China Economic Review* 27, 179–190.
- Maliszewski, W., Arslanalp, M.S., Caparusso, J., Garrido, J., Guo, M.S., Kang, J.S., Lam, W.R., Law, D., Liao, W., 2016. Resolving China's corporate debt problem. International Monetary Fund.
- Mao, W., Koo, W.W., 1997. Productivity growth, technological progress, and efficiency change in Chinese agriculture after rural economic reforms: a DEA approach. *China Economic Review* 8 (2), 157–174.
- Martins-Filho, C., Yao, F., 2015. Nonparametric stochastic frontier estimation via profile likelihood. *Econometric Reviews* 34 (4), 413–451.
- Meeusen, W., van Den Broeck, J., 1977. Efficiency estimation from Cobb-Douglas production functions with composed error. *International Economic Review* 18, 435–444.
- Modigliani, F., Miller, M.H., 1958. The cost of capital, corporation finance, and the theory of investment. *The American Economic Review* 48 (3), 261–291.

- Monchuk, D.C., Chen, Z., Bonaparte, Y., 2010. Explaining production inefficiency in China's agriculture using data envelopment analysis and semi-parametric bootstrapping. *China Economic Review* 21 (2), 346–354.
- Movshuk, O., 2004. Restructuring, productivity and technical efficiency in China's iron and steel industry, 1988–2000. *Journal of Asian Economics* 15 (1), 135–151.
- Mugera, A.W., Nyambane, G.G., 2015. Impact of debt structure on production efficiency and financial performance of broad acre farms in western Australia. *Australian Journal of Agricultural and Resource Economics* 59 (2), 208–224.
- Ng, Y.C., 2011. The productive efficiency of Chinese hospitals. *China Economic Review* 22 (3), 428–439.
- Park, B.U., Simar, L., Zelenyuk, V., 2015. Categorical data in local maximum likelihood: theory and applications to productivity analysis. *Journal of Productivity Analysis* 43, 199–214.
- Parmeter, C.F., Kumbhakar, S.C., 2014. Efficiency Analysis: a primer on recent advances. *Foundations and Trends® in Econometrics* 7 (3-4), 191–385.
- Pessarossi, P., Weill, L., 2013. Choice of corporate debt in China: the role of state ownership. *China Economic Review* 26, 1–16.
- Pettis, M., 2013. *Avoiding the fall: China's economic restructuring*. Brookings Institution Press.
- Rho, S., Schmidt, P., 2015. Are all firms inefficient? *Journal of Productivity Analysis* 43 (3), 327–349.
- Riedel, J., Jin, J., Gao, J., 2007. *How China grows: investment, finance, and reform*. Princeton University Press.
- Rosen, S., 1983. Specialization and human capital. *Journal of Labor Economics* 1 (1), 43–49.
- Shiu, A., 2002. Efficiency of Chinese enterprises. *Journal of Productivity Analysis* 18 (3), 255–267.
- Song, Z., Storesletten, K., Zilibotti, F., 2011. Growing like China. *American Economic Review* 101 (1), 196–233.
- Sun, K., Kumbhakar, S.C., 2013. Semiparametric smooth-coefficient stochastic frontier model. *Economics Letters* 120 (2), 305–309.
- Tran, K.C., Tsionas, M.G., 2015. Zero-inefficiency stochastic frontier models with varying mixing proportion: a semiparametric approach. *European Journal of Operational Research* 249 (3), 1113–1123.
- Tsai, K.H., Wang, J.C., 2004. R&D productivity and the spillover effects of high-tech industry on the traditional manufacturing sector: the case of Taiwan. *World Economy* 27 (10), 1555–1570.
- Wakelin, K., 2001. Productivity growth and R&D expenditure in UK manufacturing firms. *Research policy* 30 (7), 1079–1090.
- Wang, H.J., Schmidt, P., 2002. One-step and two-step estimation of the effects of exogenous variables on technical efficiency levels. *Journal of Productivity Analysis* 18 (2), 129–144.
- Weill, L., 2008. Leverage and corporate performance: does institutional environment matter? *Small Business Economics* 30 (3), 251–265.
- Wu, Y., 1995. The productive efficiency of Chinese iron and steel firms a stochastic frontier analysis. *Resources Policy* 21 (3), 215–222.
- Wu, Y., Zhou, X., 2013. Technical efficiency in the Chinese textile industry. *Frontiers of Economics in China* 8 (1), 146–163.
- Wu, Y., et al., 2002. Technical efficiency and its determinants in Chinese manufacturing sector. Working paper.

- Yang, J., Zeng, W., 2014. The trade-offs between efficiency and quality in the hospital production: some evidence from Shenzhen, China. *China Economic Review* 31, 166–184.
- Yao, S.J., Han, Z.W., Feng, G.F., 2007. On technical efficiency of China's insurance industry after WTO accession. *China Economic Review* 18 (1), 66–86.
- Yao, F., Zhang, F., Kumbhakar, S. C., 2018a. Semiparametric smooth coefficient stochastic frontier model with panel data. *Journal of Business & Economic Statistics*, 1–17.
- Yao, F., Wang, T., Tian, J., Kumbhakar, S. C., 2018b. Estimation of a smooth coefficient zero-inefficiency panel stochastic frontier model: a semiparametric approach. *Economics Letters* 166, 25–30.
- Zhang, F., Hall, J., Yao, F., 2018. Does economic freedom affect the production frontier? a semiparametric approach with panel data. *Economic Inquiry* 56 (2), 1380–1395.
- Zhang, R., Sun, K., Delgado, M. S., Kumbhakar, S. C., 2012. Productivity in China's high technology industry: regional heterogeneity and R&D. *Technological Forecasting and Social Change* 79 (1), 127–141.
- Zheng, J., Liu, X., Bigsten, A., 1998. Ownership structure and determinants of technical efficiency: an application of data envelopment analysis to Chinese enterprises (1986–1990). *Journal of Comparative Economics* 26 (3), 465–484.
- Zhou, X., Li, K.-W., Li, Q., 2011. An analysis on technical efficiency in post-reform China. *China Economic Review* 22 (3), 357–372.
- Zou, J., Shen, G., Gong, Y., 2018. The effect of value-added tax on leverage: Evidence from China's value-added tax reform. *China Economic Review*, In press.



**Appendix 1** Chinese Industrial Two-Digit Classification Codes

<b>CICC</b>	<b>Industry Name</b>
6	Exploration and abstention of coal and char industry
7	Exploration of oil and natural gas industry
8	Picking of ferrous metal mine industry
9	Picking of non-ferrous metal mine industry
10	Picking of nonmetal mine industry
11	Other mining industry
13	Agriculture and food processing industry
14	Foodstuff manufacturing industry
15	Soft drink manufacturing industry
16	Tobacco manufacturing industry
17	Textile industry
18	Waving costume, shoes and cap manufacturing industry
19	Leather, fur and feather manufacturing industry
20	Wood working, and wood,bamboo,bush rope,palm, and straw manufacturing industry
21	Furniture manufacturing industry
22	Paper making and paper products industry
23	Print and copy of record vehicle industry
24	Stationary and sporting goods manufacturing industry
25	Oil processing, coking and nuclear manufacturing industry
26	Chemical material and chemical product manufacturing industry
27	Medicine manufacturing industry
28	Chemical fiber manufacturing industry
29	Rubber product industry
30	Plastics product industry
31	Nonmetallic mineral product industry
32	Ferrous metal refining and calendaring processing industry
33	Non-ferrous metal refining and calendaring processing industry
34	Metal product industry
35	Universal equipment manufacturing industry
36	Task equipment manufacturing industry
37	Transport and communication facilities manufacturing industry
39	Electric machine and fittings manufacturing industry
40	Communication apparatus ,computer and other electric installation manufacturing industry
41	Instrument and meter, stationery machine manufacturing industry
42	Handicraft and other manufacturing industry
43	Removal and processing of obsolete resource and material industry
44	Electricity and thermal manufacturing and supplying industry
45	Combustion gas manufacturing and supplying industry
46	Water manufacturing and supplying industry

*Note:* CICC refers to two-digit industry classification code based on NBSC. Industries are mainly classified as manufacturing (CICC 13-43), mining (CICC 6-11), and public utility industries (CICC 44-46).

## Appendix 2 China Regional Map



## Chapter 3

# A Fixed Effect Additive Stochastic Frontier Model with Interactions and Distribution Free Inefficiency

The introduction of stochastic frontier (SF) model by Aigner et al. (1977) and Meeusen and van Den Broeck (1977) has evoked a myriad of studies to model frontier function in the presence of inefficiency.<sup>1</sup> One desirable feature of the SF model is its ability to estimate and isolate the inefficiency term from stochastic noise, which can capture unobserved output variation due to randomness (e.g., production shocks, measurement error). The SF model also allows the frontier to be properly specified, facilitating statistical inference on interested economic hypotheses.

In the literature, the SF model has been commonly specified as a parametric frontier function (e.g., Cobb-Douglas, Translog), with the inefficiency following a parametric distribution (e.g., Half-normal (Aigner et al., 1977), Exponential (Meeusen and van Den Broeck, 1977), Truncated normal (Stevenson, 1980), Gamma (Stevenson, 1980; Greene, 1980), or Uniform (Li, 1996)) and the stochastic noise following a normal distribution. The basic model setup has been further extended to incorporate panel data structure. Greene (2005a,b) consider a fixed effect SF model (SF-FE) by separating firm's unobserved heterogeneity from inefficiency, and estimate it directly with firm-specific dummies. Wang and Ho (2010) and Chen et al. (2014) estimate the SF-FE by applying within or first-differencing estimator to avoid the incidental parameter. Colombi et al. (2014) extend the approach in Chen et al. (2014) to propose a random effect SF model (SF-RE).<sup>2</sup> Nonetheless, the identification of the SF models above hinges upon correct specification on frontier model and the distribution of composite error (i.e., inefficiency and stochastic noise). Without a clear guidance from economic theories, however, the specifications are restrictive and

---

<sup>1</sup>For comprehensive reviews of SF models, see Greene (1993), Kumbhakar and Lovell (2000), Parmeter and Kumbhakar (2014), and Kumbhakar et al. (2018). For the illustration of SF empirical studies, see Kumbhakar et al. (2015). For an extensive review of Data Envelopment Analysis as an alternative approach to SF, see Simar and Wilson (2008).

<sup>2</sup>In the following discussion, we use the term SF-FE (SF-RE) to represent a generic fixed effect (random effect) SF model, regardless of the specification made on the frontier or composite error.

may not hold in practice.

Maintaining the parametric specification on the frontier model, many studies have made effort on eschewing the specification on the composite error distribution. With panel data, Schmidt and Sickles (1984) consider a time-invariant inefficiency, which can be bundled and eliminated with firm fixed effect. Cornwell et al. (1990) model the inefficiency as a deterministic function of time trend. Kumbhakar (1990), Battese and Coelli (1992), Lee and Schmidt (1993), and Kumbhakar and Wang (2005a) specify a time-invariant inefficiency, scaled by a deterministic function of time with different functional form. With cross sectional data, Horrace and Parmeter (2011) estimate the unknown distribution of the composite error through deconvolution. Parmeter et al. (2016) propose a partially linear SF model (SF-PLM) to estimate the mean function of inefficiency with distribution free composite error. However, their frontier models require a parametric specification *a priori*, which may deviate from its true underlying structure. A correctly specified frontier model is paramount to provide reliable insights on the features of technology (e.g., return to scale) as well as inefficiency level, thus rendering solid implication for policy makers.

Maintaining the parametric specification on the composite error, a series of studies have improved the flexibility of modeling the frontier. Fan et al. (1996) first propose a fully nonparametric frontier model. Martins-Filho and Yao (2015) investigate its asymptotic properties, and propose a refined profile-likelihood based estimator for the nonparametric frontier. Kumbhakar et al. (2007) model all parameters in the composite error distribution and frontier nonparametrically with kernel weighted local log-likelihood approach. Park et al. (2015) further generalize their results to incorporate categorical data. Kneip et al. (2015) propose to estimate a fully nonparametric frontier model when the stochastic noise is log-normally distributed with unknown variance.

Given many input variables in applied works, however, the fully nonparametric frontier can be empirically infeasible due to the well-known *curse of dimensionality* problem. Under the framework of Fan et al. (1996), Ferrara and Vidoli (2017) consider an additive frontier model (SF-AM) of inputs. Sun and Kumbhakar (2013) and Yao et al. (2018b) consider a semiparametric smooth coefficient model (SF-SC), where a set of environment variables can impact frontier as well as the moments of inefficiency. However, the SF-AM does not model the determinants of inefficiency. Also, both SF-AM and SF-SC cannot allow interactions among frontier inputs, which can be a common feature in practice due to the possibility of their substitutable/complementary effect. Furthermore, all non/semiparametric frontier models above do not separate time-invariant fixed effect from inefficiency. With panel data, the fixed effect may be embedded in inefficiency, and the level of which can be overestimated and thus misleading.

Hence, the restrictive assumption on composite error distribution or frontier model has been relaxed on one side at the expense of being maintained on the other side. In this paper, we relax the assumption

on both sides by proposing a fixed effect semiparametric additive frontier model with interactions, which unifies several appealing features. First, we allow a set of inputs  $x$  and environment variables  $z$  to shift the frontier through unknown smooth functions. We further incorporate interactive effects within  $x$ , within  $z$ , and between  $x$  and  $z$ , and model them as smooth functions of their multiplicative value. Hence, our model captures the potential nonlinearity of  $x$  and  $z$ , and reveals their interactive effects beyond a linear fashion. However, SF-SC cannot model the nonlinearity of inputs and the interaction within  $x$ . SF-AM and SF-PLM cannot model the interaction either within or between  $x$  and  $z$ . In fact, our model nests additive model, partially linear additive model, and linear model as special cases.

Second, we do not impose distribution assumption on both inefficiency term and stochastic error. We allow the inefficiency to be influenced by its observed determinants  $w$  through its mean function. Parmeter et al. (2016) estimate the mean function nonparametrically, which is an appealing feature. As the number of inefficiency determinants rises (see, for example, Iyer et al. (2008)), however, the nonparametric estimation is subject to the *curse of dimensionality*. To improve the model's applicability, we follow a conventional specification in the literature by assuming the mean function to be known up to certain parameters. We note that SF-AM assumes the absence of  $w$ , SF-SC require  $w$  to only include  $z$ , and the SF-PLM precludes  $w$  from having any variable in  $x$  or  $z$ . Here, we provide a more flexibility on modeling the inefficiency determinants, which can include all, some, or none of variables appearing on the frontier. As emphasized in (Parmeter and Kumbhakar, 2014), a correct modeling on the dependence of the inefficiency on  $w$  is important to avoid biasing the estimates of both frontier and technical efficiency.

Finally, we adopt a panel data structure, allowing the inefficiency to be disentangled from time-invariant fixed effect. We allow the fixed effect to be arbitrarily correlated with variables showing up on both frontier and inefficiency mean function. We note that taking into account the fixed effect is relevant to providing robust inefficiency level estimates, as well as to meaningfully mitigating the endogeneity concerns due to potential omitted variables.

We achieve the flexible modeling on the frontier with distribution free composite error through a combined use of B-spline estimator and Nonlinear Least Square estimator (NLS). Utilizing the property of B-spline estimator, we eliminate the fixed effect through within transformation, and consistently estimate unknown smooth functions as well as their derivatives. Due to the presence of inefficiency, our frontier estimates can be expressed as a function of the parameters in the inefficiency mean function, which are consistently estimated through NLS. Our estimators are free of incidental parameter and the *curse of dimensionality* problem, and computationally attractive for massive datasets in empirical application. We also propose two nonparametric hypothesis tests on the frontier model specification and the presence of interaction functions. Through simulation studies, we demonstrate the appealing finite sample

performance of our proposed estimators and tests. We apply our model and tests to a world production frontier application with 116 countries during 2000-2013.

Following the introduction from above, we detail our estimation procedure with two statistical tests in Section 2, illustrate their finite sample performance through simulation studies in Section 3, illustrate an empirical application in Section 4, and conclude in Section 5.

## 3.1 Estimation Procedure and Hypothesis Testing

### 3.1.1 Semiparametric Estimation

To facilitate our discussion, we adopt a production frontier framework, and consider the following fixed effect semiparametric additive stochastic frontier model with interactions (SF-AMIFE):

$$y_{it} = \alpha_i + \sum_{j=1}^d m_j(x_{j,it}) + \sum_{s=1}^q f_s(z_{s,it}) + \sum_{k=1}^m g_k(\zeta_{k,it}) + v_{it} - u_{it}, \quad (3.1)$$

where  $i = 1, \dots, n$  and  $t = 1, \dots, T$  indexes for firm and time, respectively,  $y_{it}$  is the production output,  $x_{it} \in \mathbb{R}^d$  is a random vector of traditional frontier inputs, such as labor and physical capital, and  $z_{it} \in \mathbb{R}^q$  represents the exogenous environment variables, such as human capital or economic freedom, having facilitating impacts on the frontier. To capture possible interaction effects, Sperlich et al. (2002) incorporate smooth bivariate functions of two different regressors in an additive model, estimated via marginal integration (Linton and Nielsen, 1995). However, the marginal integration estimator is embodied with the *curse of dimensionality*, thus impeding its practical use in empirical studies. Here, we allow the interaction to be captured through a function  $g_k(\zeta_{k,it})$ , where  $\zeta_{k,it}$  is a multiplicative value of choice from  $x_{j,it} * x_{j',it}$ ,  $z_{s,it} * z_{s',it}$ , and  $x_{j,it} * z_{s,it}$ , for  $j \neq j'$ , or  $s \neq s'$ . Hence, model (3.1) allows both  $x_{it}$  and  $z_{it}$  to shift the frontier individually through  $m_j(\cdot)$  and  $f_s(\cdot)$ , and interactively through  $g_k(\cdot)$ , all three functions being unknown and smooth. In other words, our model permits the nonlinearity of  $x$ ,  $z$ , and their multiplicative interactions, thus alleviating significantly the risk of model misspecification.

To ensure the identification of (3.1), we follow the literature to impose the condition  $E(m_j(\cdot)) = E(f_s(\cdot)) = E(g_k(\cdot)) = 0$  for all  $j = 1, \dots, d$ ,  $s = 1, \dots, q$ , and  $k = 1, \dots, m$ . The condition is clearly not a restriction in our model, which can be easily imposed in practice. Of course, the identification condition will not affect the derivative estimation of functions, which are of main interests in economic studies. Notice that when  $g_k(\cdot)$  is a linear function, (3.1) reduces to partially linear additive model (Li, 2000). When  $g_k(\cdot) = 0, \forall k = 1, \dots, m$ , (3.1) becomes purely additive model (Linton and Nielsen, 1995). When all functions  $m_j(\cdot)$ ,  $f_s(\cdot)$ , and  $g_k(\cdot)$  are linear functions, (3.1) is the conventional linear regression.

The composite error term  $\epsilon_{it} = v_{it} - u_{it}$  is consisted of the stochastic noise  $v_{it}$  capturing the randomness, and the one-sided random variable  $u_{it}$  representing inefficiency, or the shortfall of production outputs from its frontier level. We allow the conditional mean function of inefficiency to be influenced by its determinants  $w_{it} \in \mathbb{R}^p$  such that  $E(u_{it}|w_{it}) = \mu(w_{it}; \gamma)$ , with  $\mu(w_{it}; \gamma)$  a non-negative function known up to a vector of coefficients  $\gamma$ . The vector  $w_{it}$  may or may not include  $x$  and/or  $z$ . Let  $\Psi_{it} = \{x_{it}, z_{it}, w_{it}\}$ , we assume  $E(v_{it}|\Psi_{it}) = 0$ , so  $E(\epsilon_{it}|\Psi_{it}) = -\mu(w_{it}; \gamma)$ . With the use of NLS estimator illustrated below, we can eschew the distribution assumptions on both  $v_{it}$  and  $u_{it}$  for the identification of  $\gamma$ , thus circumvent the risk of distribution misspecification.

Finally,  $\alpha_i$  is the fixed effect, reflecting unobserved heterogeneities across firms that does not change over time. We assume  $\alpha_i$  is i.i.d. and  $E(\alpha_i) = 0$ . We separate it from  $u_{it}$ , and allow it to be arbitrarily correlated with  $x_{it}$ ,  $z_{it}$ , and  $w_{it}$ , so  $E(\alpha_i|\Psi_{it}) \neq 0$ . If  $E(\alpha_i|\Psi_{it}) = 0$ , model (3.1) reduces to SF-RE. It is well known, however, that the FE estimator is consistent in both SF-FE and SF-RE, except with some efficiency loss when the SF-RE model is true. However, SF-RE model is inconsistent if  $\alpha_i$  is correlated with regressors, which can commonly occur in empirical studies. Thus, we focus on the SF-FE framework in our study.

A direct estimation of (3.1) is infeasible since  $E(\epsilon_{it}|\Psi_{it}) \neq 0$ . Instead, we add  $\mu(w_{it}; \gamma)$  on both sides in (3.1) to obtain

$$\tilde{y}_{it}(\gamma) = \alpha_i + \sum_{j=1}^d m_j(x_{j,it}) + \sum_{s=1}^q f_s(z_{s,it}) + \sum_{k=1}^m g_k(\zeta_{k,it}) + \tilde{\epsilon}_{it}, \quad (3.2)$$

where  $\tilde{y}_{it}(\gamma) = y_{it} + \mu(w_{it}; \gamma)$  can be known up to  $\gamma$ , and  $\tilde{\epsilon}_{it} = \epsilon_{it} + \mu(w_{it}; \gamma)$  satisfying the  $E(\tilde{\epsilon}_{it}|\Psi_{it}) = 0$ . Hence, the frontier model in (3.2) can be identified if  $\gamma$  is known. We implement B-spline estimator to approximate unknown functions in the frontier. For a generic random variable  $\varphi_{it} \in \mathbb{R}^r$  with their smooth functions  $\{\tau_l(\varphi_{l,it})\}_{l=1}^r$ , we place  $J$  number of interior knots evenly on the support of  $\varphi_{l,it}$ , and choose the B-spline basis function with order  $\rho + 1$ . Denote  $L = J + \rho$ , we approximate  $\sum_{l=1}^r \tau_l(\varphi_{l,it}) \approx \phi_\rho(\varphi_{it})^\top \beta_\varphi$  from

$$\sum_{l=1}^r \tau_l(\varphi_{l,it}) \approx \phi_\rho(\varphi_{it})^\top \beta_\varphi, \quad (3.3)$$

where  $\phi_\rho(\varphi_{it}) = [\phi_\rho(\varphi_{1,it})^\top, \dots, \phi_\rho(\varphi_{r,it})^\top]^\top$  is a  $(1 \times rL)$  vector with

$\phi_\rho(\varphi_{l,it}) = [\phi_1(\varphi_{l,it}), \dots, \phi_L(\varphi_{l,it})]^\top$ , and  $\beta_\varphi = [\beta_{\varphi_1}^\top, \dots, \beta_{\varphi_r}^\top]^\top$  is a  $(rL \times 1)$  coefficient vector with  $\beta_{\varphi_l} = [\beta_{\varphi_{1,l}}, \dots, \beta_{\varphi_{L,l}}]^\top$ .<sup>3</sup> For each function  $\tau_l(\varphi_l)$ , we further obtain the consistent derivative estimates of  $\tau'_l(\varphi_l) \equiv \frac{\partial}{\partial \varphi_l} \tau_l(\varphi_l)$  from

$$\tau'_l(\varphi_l) \approx \phi'_\rho(\varphi_{l,it})^\top \beta_{\varphi_l}, \quad (3.4)$$

---

<sup>3</sup>Since  $\sum_{l=1}^r \phi_l(\varphi_{l,it}) = 1$ , we require  $L = J + \rho$  instead of  $L = J + \rho + 1$  due to the multicollinearity issue.

where  $\phi'_\rho(\varphi_{l,it}) = [\phi'_1(\varphi_{l,it}), \dots, \phi'_L(\varphi_{l,it})]^\top$  is a  $1 \times L$  vector, and  $\beta_{\varphi_l}$  is obtained in (3.3). We follow Ai et al. (2014) to let  $J \rightarrow \infty$  and  $J/n \rightarrow 0$  as  $n \rightarrow \infty$  for fixed  $T$ . Thus, with the B-spline estimator as in (3.3), we approximate (3.2) as

$$\tilde{y}_{it}(\gamma) = \alpha_i + \phi_\rho(x_{it}, z_{it}, \zeta_{it})^\top \beta + \tilde{\epsilon}_{it}, \quad (3.5)$$

where  $\phi_\rho(x_{it}, z_{it}, \zeta_{it})^\top = [\phi_\rho(x_{it})^\top, \phi_\rho(z_{it})^\top, \phi_\rho(\zeta_{it})^\top]^\top$  and  $\beta = [\beta_x^\top, \beta_z^\top, \beta_\zeta^\top]^\top$  are defined as in (3.3) with  $\varphi$  replaced correspondingly. To facilitate discussion, let  $\mathbf{Q} = [\mathbf{X}, \mathbf{Z}, \mathbf{\zeta}]$  be a  $nT \times (d + q + m)$  matrix with its  $it^{th}$  row given by  $Q_{it} = [x_{it}, z_{it}, \zeta_{it}]^\top$ . We rewrite (3.5) in matrix form as

$$\tilde{Y}(\gamma) = D\alpha + \Phi(\mathbf{Q})\beta + \tilde{\epsilon}, \quad (3.6)$$

where  $\tilde{Y}(\gamma) = [\tilde{y}_{11}(\gamma), \dots, \tilde{y}_{nT}(\gamma)]^\top$ ,  $\alpha = [\alpha_1, \dots, \alpha_n]^\top$ ,  $\tilde{\epsilon} = [\tilde{\epsilon}_{11}, \dots, \tilde{\epsilon}_{nT}]^\top$ ,  $\Phi(\mathbf{Q})$  is a  $nT \times L(d + q + m)$  matrix with its  $it^{th}$  row given by  $\phi_\rho(x_{it}, z_{it}, \zeta_{it})^\top$  in (3.5).  $D = I_n \otimes \iota_T$ , where  $I_e$  and  $\iota_e$  represent, respectively, a  $e \times e$  identity matrix and a  $e \times 1$  vector of ones a positive constant  $e$ . Define the orthogonal projection matrix  $M_D = I_{nT} - D(D^\top D)^{-1}D^\top$ , we wipe out the fixed effect  $\alpha$  by pre-multiplying  $M_D$  on both sides of (3.6) to have

$$M_D \tilde{Y}(\gamma) = M_D \Phi(\mathbf{Q})\beta + M_D \tilde{\epsilon}, \quad (3.7)$$

which permits the closed form solution of  $\beta$  through OLS:

$$\hat{\beta}(\gamma) = (\Phi(\mathbf{Q})^\top M_D \Phi(\mathbf{Q}))^{-1} \Phi(\mathbf{Q})^\top M_D \tilde{Y}(\gamma). \quad (3.8)$$

We then obtain the function estimates  $\hat{m}_j(\cdot)$ ,  $\hat{f}_s(\cdot)$ , and  $\hat{g}_k(\cdot)$  with (3.8) correspondingly. In practice, however,  $\hat{\beta}(\gamma)$  is infeasible since  $\gamma$  is unknown. We propose to replace  $\gamma$  by its consistent NLS estimates  $\hat{\gamma}$  from

$$\hat{\gamma} = \underset{\gamma}{\operatorname{argmin}} [\tilde{Y}(\gamma) - \Phi(\mathbf{Q})\beta(\gamma)]^\top M_D [\tilde{Y}(\gamma) - \Phi(\mathbf{Q})\beta(\gamma)]. \quad (3.9)$$

With  $\hat{\gamma}$ , we update our B-spline estimator  $\hat{\beta}(\gamma)$  in (3.8) by  $\hat{\beta}(\hat{\gamma})$ , as well as the estimates of inefficiency mean function by  $\mu(w_{it}; \hat{\gamma})$ . Finally, we obtain the estimates of the partial effect of  $x_j$  and  $z$  on the frontier for a fixed  $i$  and  $t$ , respectively, from

$$\frac{\partial y}{\partial x_j} = \phi'_\rho(x_j)^\top \hat{\beta}_{x_j}(\hat{\gamma}) + \sum_{j'} \phi'_\rho(x_j x_{j'})^\top \beta_{\zeta_{jj'}}(\hat{\gamma}) x_{j'} + \sum_s \phi'_\rho(x_j z_s)^\top \beta_{\zeta_{js}}(\hat{\gamma}) z_s, \quad (3.10)$$



and

$$\frac{\partial y}{\partial z_s} = \phi'_\rho(z_s)^\top \hat{\beta}_{z_s}(\hat{\gamma}) + \sum_{s'} \phi'_\rho(z_s z_{s'})^\top \beta_{\zeta_{ss'}}(\hat{\gamma}) z_{s'} + \sum_j \phi'_\rho(x_j z_s)^\top \beta_{\zeta_{js}}(\hat{\gamma}) x_j, \quad (3.11)$$

where the  $\phi'_\rho(\cdot)$  is define in (3.4).

### 3.1.2 Hypothesis Testing

We propose two hypothesis tests that are relevant to empirical application. First, it is important to provide statistical foundation for the functional form of SF-SIMFE in (3.1). Thus, one may interested in testing whether (3.1) can be correctly specified by parametric functions. If so, a parametric SF model should be implemented for significant estimation efficiency gain. Denote  $\Lambda(Q) = \sum_{j=1}^d m_j(x_j) + \sum_{s=1}^q f_s(z_s) + \sum_{k=1}^m g_k(\zeta_k)$  defined in (3.2). Let  $\Lambda(Q; \eta) = \sum_{j=1}^d m_j(x_j; \eta_x) + \sum_{s=1}^q f_s(z_s; \eta_s) + \sum_{k=1}^m g_k(\zeta_k; \eta_\zeta)$ , where  $m_j(x_j; \eta_x)$ ,  $f_s(z_s; \eta_s)$ , and  $g_k(\zeta_k; \eta_\zeta)$  be parametric functions known up to a finite vector of  $\eta = (\eta_x, \eta_z, \eta_\zeta)$ . We are interested in testing the following hypothesis on the parametric specification of the frontier model:

$$H_{01} : \Pr\{\Lambda(Q) = \Lambda(Q; \eta)\} = 1, \quad (3.12)$$

for almost all  $x$ ,  $z$ , and  $\zeta$  in  $Q$ , and some  $\eta$  in a compact subset of  $\mathbb{R}^{d+q+m}$ . Assuming that  $\hat{\gamma}_0$  is the  $\sqrt{n}$ -consistent estimator for  $\gamma_0$  under  $H_{01}$ , we observe from (3.2) that

$$\tilde{y}_{it}(\gamma_0)(1 + o_p(1)) - \Lambda(Q_{it}; \eta) = \alpha_i + \tilde{\epsilon}_{it}. \quad (3.13)$$

For simplicity without losing generosity, we consider a linear parametric model as a special case of  $\Lambda(Q_{it}; \eta)$ . Let  $\Lambda(\mathbf{Q}; \eta)$  be the  $nT \times 1$  matrix with the  $it^{th}$  row equal to  $\Lambda(Q_{it}; \eta)$ , we have  $\Lambda(\mathbf{Q}; \eta) \equiv \mathbf{Q}\eta(\gamma_0)$ , where  $\eta(\gamma_0) = [\mathbf{Q}^\top M_D \mathbf{Q}]^{-1} \mathbf{Q}^\top M_D \tilde{Y}(\gamma_0)$ . Thus, under  $H_{01}$  we observe in (3.13) that  $M_D(\tilde{Y}(\gamma_0) - \mathbf{Q}\eta) = M_D \tilde{\epsilon} \equiv e$ . Inspired by Lin et al. (2014), we focus on a conditional moment test  $I_{n1} = E(eE(e|x, z)p(x, z))$ , with  $p(x, z)$  the multivariate density of  $x$  and  $z$ . Clearly,  $I_{n1} = 0$  under  $H_{01}$  since  $E(e|x, z) = 0$ . We construct the following feasible test statistic of  $I_{n1}$  with local constant estimator of  $E(e|x, z)$ :

$$\hat{I}_{n1} = \frac{1}{n^2 h^{d+q}} \sum_{i=1}^n \sum_{t=1}^n \sum_{i' \neq i}^n \sum_{t' \neq t}^T \hat{e}_{it} \hat{e}_{i't'} K\left(\frac{x_{it} - x_{i't'}}{h_x}\right) K\left(\frac{z_{it} - z_{i't'}}{h_z}\right), \quad (3.14)$$

with its standardized version given by

$$\hat{T}_{n1} = \frac{n\sqrt{h^{h+q}} \hat{I}_{n(1)}}{\sqrt{\hat{\Sigma}_n}}, \quad (3.15)$$

where  $\hat{\Sigma}_n = \frac{2}{n^2 h^{d+q}} \sum_{i=1}^n \sum_{t=1}^n \sum_{i' \neq i}^n \sum_{t' \neq t}^T \hat{e}_{it}^2 \hat{e}_{i't'}^2 K^2\left(\frac{x_{it} - x_{i't'}}{h_x}\right) K^2\left(\frac{z_{it} - z_{i't'}}{h_z}\right)$ , and  $K(\cdot) : \mathbb{R}^r \rightarrow \mathbb{R}$  with  $r \in (d, q)$  is product kernel function.  $\hat{e}_{it}$  is the  $it^{th}$  element of  $\hat{e} = M_D(\tilde{Y}(\hat{\gamma}_0) - \mathbf{Q}\hat{\eta}(\hat{\gamma}_0))$ , where  $\hat{\eta}(\hat{\gamma}_0) = [\mathbf{Q}^\top M_D \mathbf{Q}]^{-1} \mathbf{Q}^\top M_D \tilde{Y}(\hat{\gamma}_0)$ . Given the condition  $\hat{\gamma}_0 - \gamma_0 = o_p(1/\sqrt{n})$ , we note that tests  $\hat{T}_{n1}$  are the same as the test in equation (3.5) of Lin et al. (2014). Thus, we expect that  $\hat{T}_{n1} \xrightarrow{d} N(0, 1)$  when  $H_{01}$  is true.

In addition to the functional form test, one may also interested in testing possible interaction effects between inputs and environment variables of interests. For instance, the output elasticities of capital and labor can significantly depend on the human capital (Yao et al., 2018b) or economic freedom (Zhang et al., 2018). Since we capture potential interactive effects through  $g_k(\cdot)$ , providing statistical evidence on whether  $g_k(\cdot)$  appears on the frontier model can be very useful to support/refute its related economic hypotheses. Specifically, we test the following null hypothesis:

$$H_{02} : \Pr\{g_k(\zeta_k) = 0\} = 1, \quad (3.16)$$

for almost all  $\zeta_k$  defined in (3.1), and  $k = 1, \dots, m$ . Inspired by Sperlich et al. (2002), we consider a L-2 norm based test statistic  $I_{n2} = \int_{\mathbb{R}} g_k^2(\zeta) p(\zeta) d\zeta$ , with  $p(\zeta)$  the density of  $\zeta$ .  $I_{n2}$  is a suitable candidate for testing  $H_{02}$ , since  $I_{n2} = 0$  if and only if  $H_{02}$  is true. In practice, we use the sample analogue of  $I_{n2}$  to form a feasible test as:

$$\hat{I}_{n2} = \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^n \hat{g}_k(\zeta_{k,it})^2 w(\zeta_{k,it}), \quad (3.17)$$

where  $\hat{g}_k(\zeta_{k,it})$  is obtained correspondingly from (3.8), and  $w(\cdot)$  is a trimming function to remove outliers of  $\zeta$  due to the possible singularity of the inverse matrix in (3.8).

It is well known that the size and power performance based on the asymptotic null distribution of  $\hat{T}_{n1}$  or  $\hat{I}_{n2}$  can be poor in practice unless the sample size can be sufficiently large. Thus, we approximate the distribution of  $\hat{T}_{n1}$  ( $\hat{I}_{n2}$ ) under  $H_{01}$  ( $H_{02}$ ) through a wild-bootstrapped procedure as follows:

- 1) With the original sample  $\Theta = \{y_{it}, Q_{it}\}_{i=1, t=1}^{n, T}$ , compute  $\hat{T}_{n1}$  in (3.15) or  $\hat{I}_{n2}$  in (3.17). Obtain the residual  $\hat{e}_{it} = \hat{\epsilon}_{it} - \frac{1}{T} \sum_{t=1}^T \hat{\epsilon}_{it}$ , where  $\hat{\epsilon}_{it} = \tilde{y}_{it}(\hat{\gamma}) - \hat{\Lambda}(Q_{it})$ , with  $\hat{\Lambda}(Q_{it}) = \phi_\rho(Q_{it})^\top \hat{\beta}(\hat{\gamma})$  defined in (3.6).
- 2) With  $\{\hat{e}_{it}\}_{i=1, t=1}^{n, T}$ , obtain wild-bootstrap residual  $\{e_{it}^*\}_{i=1, t=1}^{n, T}$ , where  $e_{it}^* = [(1 - \sqrt{5})/2] \hat{e}_{it}$  with probability  $P = (1 + \sqrt{5})/2\sqrt{5}$ , and  $e_{it}^* = [(1 + \sqrt{5})/2] \hat{e}_{it}$  with probability  $1 - P$ .

For testing  $H_{01}$ , construct the bootstrap sample  $\Theta_1^* = \{y_{it}^*, Q_{it}\}_{i=1, t=1}^{n, T}$ , where  $y_{it}^* = \hat{\Lambda}(Q_{it}; \hat{\gamma}_0^*) + e_{it}^*$ , with  $\hat{\Lambda}(Q_{it}; \hat{\gamma}_0^*) = Q_{it}^\top \hat{\eta}(\hat{\gamma}_0^*)$ , and  $\hat{\eta}(\hat{\gamma}_0^*) = [\mathbf{Q}^\top M_D \mathbf{Q}]^{-1} \mathbf{Q}^\top M_D \tilde{Y}^*(\hat{\gamma}_0^*)$ .

For testing  $H_{02}$ , construct the bootstrap sample  $\Theta_2^* = \{y_{it}^*, x_{it}, z_{it}, \zeta_{it}\}_{i=1, t=1}^{n, T}$ , where  $y_{it}^* = \hat{\Lambda}_0(Q_{-\zeta_{k,it}}) + e_{it}^*$ , with  $\hat{\Lambda}_0(Q_{-\zeta_{k,it}}) \equiv \phi_\rho(Q_{-\zeta_{k,it}})^\top \hat{\beta}_0(\hat{\gamma}_0)$ , and  $Q_{-\zeta_{k,it}}$  is the vector with the  $k^{th}$  element in  $\zeta$  re-

moved from  $Q_{it}$ .

- 3) Compute  $\hat{T}_{n1}^*$  ( $\hat{I}_{n2}^*$ ) similar as in 1), except replacing  $\Theta$  with  $\Theta_1^*$  ( $\Theta_2^*$ ) for  $H_{01}$  ( $H_{02}$ ).
- 4) Repeat step 2)-3) a large number (B) of times to obtain an empirical distribution from  $\{\hat{T}_{n1,b}^*\}_{b=1}^B$  or  $\{\hat{I}_{n2,b}^*\}_{b=1}^B$ . We reject  $H_{01}$  ( $H_{02}$ ) if  $p_{n1}^*(p_{n2}^*) < a$ , where  $p_{n1}^* = \frac{1}{B} \sum_{b=1}^B \mathbf{1}(\hat{T}_{n1,b}^* > \hat{T}_{n1})$  ( $p_{n2}^* = \frac{1}{B} \sum_{b=1}^B \mathbf{1}(\hat{I}_{n2,b}^* > \hat{I}_{n2})$ ) is the empirical p-value,  $\mathbf{1}(\cdot)$  is the indicator function, and  $a$  is the significant level.

In each repetition above, we generate the bootstrapped dependent variable  $y_{it}^*$  correspondingly to impose the null  $H_{01}$  or  $H_{02}$ . Hence, the empirical distribution from  $\{\hat{T}_{n1,b}^*\}_{b=1}^B$  ( $\{\hat{I}_{n2,b}^*\}_{b=1}^B$ ) can be used to approximate the null distribution of  $\hat{T}_{n1}$  ( $\hat{I}_{n2}$ ), regardless of whether  $H_{01}$  ( $H_{02}$ ) holds or not. For example, if  $H_{01}$  is true, the bootstrap procedure can asymptotically lead to the correct size of the  $\hat{T}_{n1}$ . Thus, we expect that  $\sup_{\xi \in \mathbb{R}} |\Pr(\hat{T}_{n(1)}^* < \xi | \Theta) - \Phi(\xi)| = o_p(1)$ , where  $\Phi(\cdot)$  is the standard normal c.d.f.. If  $H_{01}$  is not true, our bootstrap critical value remains asymptotically finite for any  $a > 0$ , but  $\hat{T}_{n1}$  will converge to infinity. Thus, we can show that  $P(\hat{T}_{n1} > \hat{T}_{n1}^* | H_{01} \text{ is false}) \rightarrow 1$ , indicating a consistent bootstrap test. We expect that the similar argument can be applied to the asymptotic property of  $\hat{I}_{n2}$  and its standardized version.

## 3.2 Monte Carlo Simulation

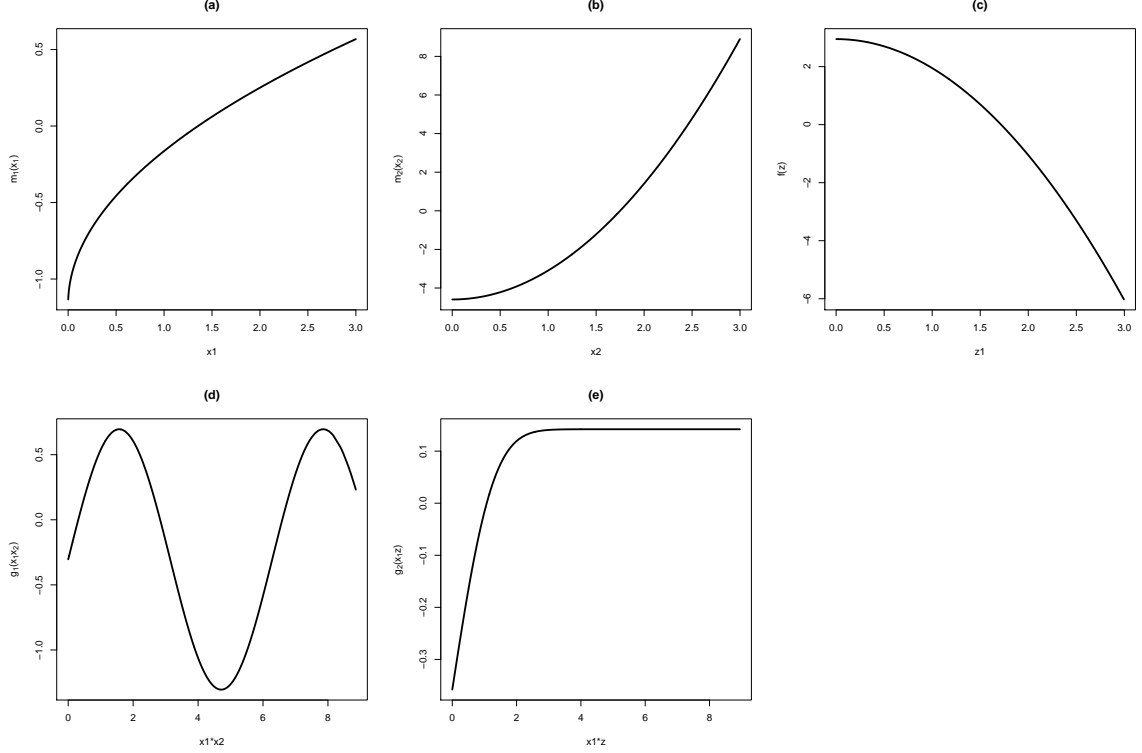
In this section, we investigate the finite sample performance of our estimators of frontier and parameters in the inefficiency mean function, and of two tests  $\hat{T}_{n1}$  and  $\hat{I}_{n2}$  proposed in Section 2. We first explore the performance of  $\hat{\gamma}$  in (3.9) and  $\hat{\beta}(\hat{\gamma})$  in (3.8) under 1) sample size  $(n, T)$  with different combination, 2) SF-RE or SF-FE model, and 3) correlated regressors between inputs and environment variables.

We consider the following Data Generating Process (DGP) for the simulation study:

$$y_{it} = \alpha_i + m_1(x_{1,it}) + m_2(x_{2,it}) + f(z_{it}) + g_1(x_{1,it}x_{2,it}) + g_2(x_{1,it}z_{it}) + v_{it} - u_{it}, \quad (3.18)$$

where for  $i = 1, \dots, n$  and  $t = 1, \dots, T$ ,  $m_1(x_{1,it}) = \sqrt{x_{1,it}}$ ,  $m_2(x_{2,it}) = \frac{3}{2}x_{2,it}^2$ ,  $f(z_{it}) = -z_{it}^2$ ,  $g_1(x_{1,it}x_{2,it}) = \sin(x_{1,it}x_{2,it})$ , and  $g_2(x_{1,it}z_{it}) = \Phi(x_{1,it}z_{it})$ , with  $\Phi$  the standard normal c.d.f.. To mimic the possible correlation across frontier variables, we generate variables from a multivariate normal distribution  $\{x_{1,it}, x_{2,it}, z_{it}\}_{i=1, t=1}^{n, T} \sim N(\mu_0, \Sigma_0)$ , where  $\mu_0 = [4, 6, 2]$  is the vector of variables mean, and  $\Sigma_0$  the covariance matrix with its  $(\mathbf{i}, \mathbf{j})^{th}$  element  $\Sigma_{\mathbf{ij}} = \delta^{|\mathbf{i}-\mathbf{j}|}$  for  $\mathbf{i} = \mathbf{j} = 1, 2, 3$ . We set  $\delta = 0.5$  to allow for regressors to be fairly correlated, which can be a common feature in economic studies. To facilitate placing knots, we rescale  $\{x_{1,it}, x_{2,it}, z_{it}\}_{i=1, t=1}^{n, T}$  into a range of  $[0, 3]$ . We consider a scaling property

Figure 3.1: Simulation Function



in  $u_{it}$  such that  $u_{it} = (\sqrt{2\sigma_u^2/\pi})^{-1} u_i^* \mu(w_{it}; \gamma)$ , where  $u_i^* \sim |N(0, \sigma_u^2)|$  follows a half-normal distribution such that  $E(u_i^* | w_{it}) = \sqrt{\frac{2\sigma_u^2}{\pi}}$ . We draw  $v_{it} \sim N(0, \sigma_v^2)$ , and fix  $(\sigma_v^2, \sigma_u^2) = (0.5, 1)$  to reflect a relatively large variation of inefficiency in the composite error. Since  $u_{it} \geq 0$  by construction, we impose the non-negativity constrain by assuming  $\mu(\cdot; \gamma)$  be an exponential function. Thus, our inefficiency mean function gives  $E(u_{it} | w_{it}) \equiv \mu(w_{it}; \gamma) = \exp\{w_{1,it}\gamma_1 + z_{it}\gamma_2\}$ , where  $w_{1,it} \sim U(0, 4)$  represents a determinant of inefficiency not appearing on the frontier, and  $(\gamma_1, \gamma_2) = (-0.5, 1.5)$ . We construct the fixed effect  $\alpha_i = \frac{1}{T} \sum_{t=1}^n c_0(x_{1,it} + x_{2,it} + z_{it} + w_{1,it}) + \xi_i$ , where  $\xi_i \sim N(0, 1)$  and  $c_0$  is a constant. Thus, we set  $c_0 = 0$  ( $c_0 = 1$ ) to reflect a SF-RE (SF-FE) model. We also replace  $\alpha_i$  with  $(\alpha_i - \frac{1}{n} \sum_{i=1}^n \alpha_i)$  due to  $E(\alpha_i) = 0$ . Finally, we impose the identification condition in SF-AMIFE by empirically centering all smooth functions in (3.18). Figure 3.1 plots the simulated functions ( $m_1(\cdot), m_2(\cdot), f(\cdot)$ ) in panel (a)-(c), and interaction functions ( $g_1(\cdot), g_2(\cdot)$ ) in panel (d)-(f), respectively. Thus, the DGP displays clear nonlinearities of inputs and environment variables, with their interactive effects enter the frontier model beyond a simple linear fashion.

As a common feature in other smoothing estimators, it is important to choose properly the number of interior knots  $J$ . In the simulation study, Wang and Xue (2015) estimate an additive model with regressor defined in range  $[0, 1]$ , and choose  $J$  as the integer part of  $(n)^{\frac{1}{2\rho+3}}$ , where  $\rho$  is the polynomial

order of B-spline basis function in (3.3). To alleviate computational burden, we adopt their knots selection procedure with some modification. For neutral functions ( $m_1(\cdot)$ ,  $m_2(\cdot)$ ,  $f(\cdot)$ ) in (3.18), we choose  $J$  as the three-fold integer part of  $(nT)^{\frac{1}{2\rho+3}}$ , since  $(x_1, x_2, z_1)$  are defined in  $[0, 3]$ . For interaction functions ( $g_1(\cdot)$ ,  $g_2(\cdot)$ ), we choose  $J$  be the nine-fold integer part of  $(nT)^{\frac{1}{2\rho+3}}$ , since  $x_1x_2$  and  $x_1z_1$  has its support in  $[0, 9]$ .

For the simulation study, we follow the literature to employ a cubic B-spline basis function ( $\rho = 3$ ). We choose  $n$  from (100, 200, 400),  $T$  from (3, 4, 5), and perform  $R = 1000$  repetition in our experiment. To evaluate the performance of parameters  $\hat{\gamma} = (\hat{\gamma}_1, \hat{\gamma}_2)$ , we report the root mean squared error ( $\frac{1}{R} \sum_{r=1}^R \sqrt{(\hat{\gamma}_{p,r} - \gamma_p)^2}$ , or RMSE), the bias ( $\frac{1}{R} \sum_{r=1}^R (\hat{\gamma}_{p,r} - \gamma_p)$ , or BIAS), and the standard deviation ( $sd(\{\hat{\gamma}_p\}_{r=1}^R)$ , or STD), for  $p \in (1, 2)$ . To evaluate the performance of function estimates ( $m_1(\cdot)$ ,  $m_2(\cdot)$ ,  $f(\cdot)$ ,  $g_1(\cdot)$ ,  $g_2(\cdot)$ ), we report for each function, say  $m_1(\cdot)$ , by

- 1) RAMSE, or the root averaged MSE:  $\frac{1}{R} \sum_{r=1}^R \left[ \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^n (\hat{m}_{1r}(x_{1,it}) - m_1(x_{1,it}))^2 \right]^{\frac{1}{2}}$
- 2) ABIAS, or the averaged bias:  $\frac{1}{R} \sum_{r=1}^R \left[ \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^n (\hat{m}_{1r}(x_{1,it}) - m_1(x_{1,it})) \right]$ , and
- 3) ASTD, or the averaged standard deviation:  $\frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^n sd(\hat{m}_{1r}(x_{1,it}), \dots, \hat{m}_{1R}(x_{1,it}))$ .

We first report the simulation results of  $\hat{\gamma}$  in Table 3.1, where *Panel A* displays the results for a SF-RE model ( $c_0 = 0$ ), and *Panel B* a SF-FE model ( $c_0 = 1$ ). For comparison purpose, in each panel we compare the performance of our estimator in (3.8), denoted as FE, with our pool estimator that ignores the panel structure, denoted as Pool.<sup>4</sup> In each panel, we record the measures for FE and Pool by having  $n = (100, 200, 400)$  with  $T$  fixed at 3, 4, or 5. When the true model is SF-RE, we observe that our proposed FE estimates perform fairly well as sample size  $nT$  rises, represented by a clear declining trend toward zero in RMSE, BIAS, and STD. This is true when either  $n$  doubles with  $T$  fixed or  $T$  rises with  $n$  fixed at each level. Thus, our proposed estimator in (3.9) is unbiased and consistent. The Pool estimator is also consistent with large sample size, but has its statistical measures uniformly higher than that of FE by a large magnitude. As expected, ignoring the random effect structure can lead to an efficiency loss in the Pool estimator. When the true model is fixed effect, the performance of FE remains fairly well, again demonstrating its consistency. In contrast, the Pool estimator is clearly biased and inconsistent as expected.

We now evaluate the simulation results of function estimates in Table 3.2 for SF-RE, and in Table 3.3 for SF-FE. In each table, we report the results with  $n = (100, 200, 400)$  when  $T = 3$  in *Panel A*,  $T = 4$  in *Panel B*, and  $T = 5$  in *Panel C*. We observe that the performance between FE and Pool is similar to those in Table 3.1. Under both random and fixed effect panel data structure, our FE estimator is consistent, with RAMSE, ABIAS, and STD significantly lower than its Pool estimator counterparts (see

<sup>4</sup>In this case, the Pool estimator is obtained by replacing the matrix  $M_D\Phi(Q)$  with  $\Phi(Q)$  in (3.8) and (3.9).

Table 3.1: Simulation Results of Parameters  $\hat{\gamma}$  in Inefficiency Mean Function

<i>Panel A</i>		n=100		200		400	
$c_0 = 0$	$(T = 3)$	FE	Pool	FE	Pool	FE	Pool
$\hat{\gamma}_1$	RMSE	0.3940	1.6169	0.2003	0.8001	0.0646	0.4822
	ABIAS	-0.1174	-0.4438	-0.0315	-0.2477	-0.0101	-0.0701
	ASD	0.3762	1.5556	0.2004	0.7731	0.0638	0.3759
$\hat{\gamma}_2$	RMSE	0.3965	1.6250	0.2048	0.8042	0.0747	0.4890
	ABIAS	-0.1144	-0.4435	-0.0255	-0.4528	-0.0067	-0.0636
	ASD	0.3794	1.5571	0.2064	0.7817	0.0686	0.3798
$(T = 4)$							
$\hat{\gamma}_1$	RMSE	0.2444	1.4181	0.1415	0.6141	0.0511	0.2664
	ABIAS	-0.0495	-0.4182	-0.0195	-0.1389	-0.0060	-0.0430
	ASD	0.2395	1.2691	0.1402	0.5985	0.0508	0.2630
$\hat{\gamma}_2$	RMSE	0.2502	1.3297	0.1476	0.6254	0.0530	0.2682
	ABIAS	-0.0420	-0.3566	-0.0104	-0.1359	0.0029	-0.0316
	ASD	0.2446	1.3815	0.1481	0.6097	0.0614	0.2746
$(T = 5)$							
$\hat{\gamma}_1$	RMSE	0.2039	1.2653	0.0814	0.5202	0.0442	0.2013
	ABIAS	-0.0337	-0.4016	-0.0131	-0.1039	-0.0050	-0.0231
	ASD	0.2012	0.9387	0.0804	0.5100	0.0440	0.2001
$\hat{\gamma}_2$	RMSE	0.2092	0.9660	0.0876	0.5213	0.0440	0.2036
	ABIAS	-0.0280	-0.3954	-0.0075	-0.1023	-0.0013	-0.0180
	ASD	0.2082	1.0446	0.0825	0.5110	0.0524	0.2042
<i>Panel B</i>		n=100		200		400	
$c_0 = 1$	$(T = 3)$	FE	Pool	FE	Pool	FE	Pool
$\hat{\gamma}_1$	RMSE	0.3813	1.7057	0.2177	1.7623	0.0820	1.8668
	ABIAS	-0.1040	-0.4835	-0.0414	-0.5190	-0.0091	-0.6881
	ASD	0.3671	1.6366	0.2138	1.3053	0.0721	1.4544
$\hat{\gamma}_2$	RMSE	0.3827	1.7179	0.2271	1.7695	0.0632	1.4681
	ABIAS	-0.0962	-0.4785	-0.0400	-0.4168	-0.0009	-0.4967
	ASD	0.3696	1.6460	0.2156	1.7423	0.0616	0.9577
$(T = 4)$							
$\hat{\gamma}_1$	RMSE	0.2659	1.2639	0.0918	1.6793	0.0529	1.6116
	ABIAS	-0.0571	-0.3444	-0.0113	-0.4722	-0.0063	-0.4748
	ASD	0.2599	1.2167	0.0911	1.1575	0.0526	1.3069
$\hat{\gamma}_2$	RMSE	0.2729	1.2663	0.1000	1.3891	0.0601	1.4617
	ABIAS	-0.0450	-0.3366	-0.0091	-0.3661	-0.0006	-0.4436
	ASD	0.2634	1.2282	0.0973	1.2654	0.0539	1.2167
$(T = 5)$							
$\hat{\gamma}_1$	RMSE	0.2316	1.3015	0.0830	1.5537	0.0449	1.6007
	ABIAS	-0.0451	-0.3035	-0.0093	-0.3524	-0.0054	-0.4937
	ASD	0.2273	0.9549	0.0724	0.9526	0.0446	1.0994
$\hat{\gamma}_2$	RMSE	0.2379	1.0094	0.0744	1.5601	0.0549	1.5096
	ABIAS	-0.0367	-0.2932	-0.0010	-0.3514	-0.0003	-0.3401
	ASD	0.2395	0.9601	0.0780	0.9707	0.0513	0.9108

Table 3.2: Simulation Results of Function Estimates: A RE model ( $c_0 = 0$ )

<i>Panel A</i>		$n = 100$		200		400	
	$(T = 3)$	FE	Pool	FE	Pool	FE	Pool
$\hat{m}_1(x_1)$	RAMSE	0.2860	0.7573	0.1727	0.7572	0.1231	0.7746
	ABIAS	-0.1724	-0.5377	-0.1156	-0.5503	-0.0820	-0.5731
	ASD	0.5439	0.4450	0.4883	0.4122	0.4481	0.4115
$\hat{m}_2(x_2)$	RAMSE	0.2466	0.5111	0.1281	0.4342	0.0910	0.4156
	ABIAS	0.1371	0.3282	0.0875	0.2888	0.0621	0.2866
	ASD	0.4060	3.7864	0.4053	3.7679	0.4043	3.7554
$\hat{f}(z)$	RAMSE	0.9190	13.1011	0.6080	10.2681	0.3834	7.4158
	ABIAS	-0.5474	-8.2243	-0.3502	-6.2666	-0.2324	-4.3472
	ASD	0.9978	13.6577	0.6887	10.9940	0.4835	8.1376
$\hat{g}_1(x_1x_2)$	RAMSE	0.3398	0.6901	0.1767	0.5821	0.1273	0.5586
	ABIAS	-0.1911	-0.4619	-0.1217	-0.4172	-0.0867	-0.4190
	ASD	0.7012	0.8545	0.6399	0.7691	0.6238	0.7450
$\hat{g}_2(x_1z)$	RAMSE	0.2603	0.6697	0.1692	0.6149	0.1209	0.5886
	ABIAS	0.1668	0.4603	0.1167	0.4433	0.0831	0.4454
	ASD	0.2860	0.7948	0.2207	0.7588	0.2009	0.7423
<i>Panel B</i>		$n = 100$		200		400	
	$(T = 4)$	FE	Pool	FE	Pool	FE	Pool
$\hat{m}_1(x_1)$	RAMSE	0.2319	0.7555	0.1447	0.7720	0.0968	0.7801
	ABIAS	-0.1422	-0.5441	-0.0964	-0.5659	-0.0648	-0.5793
	ASD	0.5141	0.4248	0.4737	0.4183	0.4375	0.4133
$\hat{m}_2(x_2)$	RAMSE	0.1851	0.4648	0.1075	0.4429	0.0729	0.4177
	ABIAS	0.1073	0.3021	0.0726	0.3017	0.0500	0.2918
	ASD	0.4050	3.7728	0.4046	3.7457	0.4034	3.7416
$\hat{f}(z)$	RAMSE	0.6857	12.1005	0.4977	9.1484	0.3120	6.2823
	ABIAS	-0.3994	-7.5012	-0.2957	-5.5278	-0.1881	-3.6684
	ASD	0.7576	12.7541	0.5761	9.7940	0.4208	7.0800
$\hat{g}_1(x_1x_2)$	RAMSE	0.2581	0.6222	0.1483	0.5965	0.1025	0.5632
	ABIAS	-0.1497	-0.4264	-0.1017	-0.4385	-0.0701	-0.4308
	ASD	0.6645	0.7980	0.6324	0.7718	0.6178	0.7471
$\hat{g}_2(x_1z)$	RAMSE	0.2063	0.6410	0.1400	0.5802	0.0970	0.5738
	ABIAS	0.1393	0.4532	0.0970	0.4267	0.0667	0.4402
	ASD	0.2441	0.7778	0.1989	0.7284	0.1882	0.7297
<i>Panel C</i>		$n = 100$		200		400	
	$(T = 5)$	FE	Pool	FE	Pool	FE	Pool
$\hat{m}_1(x_1)$	RAMSE	0.1763	0.7575	0.1247	0.7721	0.0863	0.7841
	ABIAS	-0.1184	-0.5476	-0.0825	-0.5694	-0.0578	-0.5843
	ASD	0.4871	0.4188	0.4549	0.4127	0.4355	0.4145
$\hat{m}_2(x_2)$	RAMSE	0.1345	0.4460	0.0918	0.4190	0.0641	0.4156
	ABIAS	0.0916	0.2984	0.0626	0.2862	0.0439	0.2948
	ASD	0.4043	3.7630	0.4038	3.7527	0.4029	3.7374
$\hat{f}(z)$	RAMSE	0.5945	10.6131	0.3989	8.1730	0.2759	5.6196
	ABIAS	-0.3498	-6.5784	-0.2372	-8.604	-0.1661	-3.3783
	ASD	0.6654	11.2398	0.5012	8.8696	0.3945	6.3367
$\hat{g}_1(x_1x_2)$	RAMSE	0.1842	0.6005	0.1264	0.5614	0.0898	0.5609
	ABIAS	0.1269	0.4230	0.0872	0.4145	0.0618	0.4345
	ASD	0.6425	0.7791	0.6249	0.7480	0.6157	0.7418
$\hat{g}_2(x_1z)$	RAMSE	0.1692	0.6187	0.1256	0.5978	0.0879	0.5711
	ABIAS	0.1146	0.4403	0.0853	0.4501	0.0603	0.4419
	ASD	0.2265	0.7592	0.1606	0.7503	0.1044	0.7283

Table 3.3: Simulation Results of Function Estimates: A FE model ( $c_0 = 1$ )

<i>Panel A</i>		$n = 100$		200		400	
	$(T = 3)$	FE	Pool	FE	Pool	FE	Pool
$\hat{m}_1(x_1)$	RAMSE	0.2813	0.8884	0.1712	0.8749	0.1205	0.8854
	ABIAS	-0.1727	-0.6011	-0.1145	-0.5987	-0.0804	-0.6127
	ASD	0.5369	0.6018	0.4822	0.5593	0.4577	0.5559
$\hat{m}_2(x_2)$	RAMSE	0.2284	0.5362	0.1253	0.4390	0.0901	0.4058
	ABIAS	0.1336	0.3437	0.0864	0.2756	0.0614	0.2513
	ASD	0.4063	3.9534	0.4051	3.9160	0.4043	3.7374
$\hat{f}(z)$	RAMSE	0.8999	14.5179	0.6086	11.0904	0.3999	8.5688
	ABIAS	-0.5342	-9.4453	-0.3475	-6.7981	-0.2355	-5.1438
	ASD	0.9666	14.8980	0.6889	11.6298	0.5017	9.2296
$\hat{g}_1(x_1x_2)$	RAMSE	0.3087	0.9197	0.1713	0.8241	0.1250	0.8159
	ABIAS	-0.1848	-0.6259	-0.1186	-0.5996	-0.0859	-0.6246
	ASD	0.6909	1.0226	0.6391	0.9337	0.6248	0.9207
$\hat{g}_2(x_1z)$	RAMSE	0.2570	0.9655	0.1710	0.8752	0.1201	0.823
	ABIAS	0.1743	0.6718	0.1172	0.6448	0.0830	0.6311
	ASD	0.2980	1.0978	0.2286	1.0228	0.1913	0.9781
<i>Panel B</i>		$n = 100$		200		400	
	$(T = 4)$	FE	Pool	FE	Pool	FE	Pool
$\hat{m}_1(x_1)$	AMSE	0.2109	0.8443	0.1425	0.8549	0.0982	0.8761
	ABIAS	-0.1417	-0.5805	-0.0949	-0.5956	-0.0655	-0.6198
	ASD	0.5089	0.5361	0.4703	0.5254	0.4400	0.5316
$\hat{m}_2(x_2)$	AMSE	0.1566	0.4781	0.1046	0.4202	0.0738	0.4032
	ABIAS	0.1065	0.3017	0.0719	0.2628	0.0500	0.2538
	ASD	0.4057	3.8879	0.4045	3.8652	0.4038	3.8452
$\hat{f}(z)$	RMSE	0.7023	12.5170	0.4681	9.9073	0.3252	7.0054
	ABIAS	-0.4148	-7.8225	-0.2806	-6.0209	-0.1987	-4.2254
	ASD	0.7769	13.0300	0.5452	10.4745	0.4300	7.7369
$\hat{g}_1(x_1x_2)$	AMSE	0.2172	0.8161	0.1452	0.7620	0.1024	0.7617
	ABIAS	-0.1483	-0.5753	-0.1006	-0.5672	-0.0701	-0.5899
	ASD	0.6546	0.9355	0.6290	0.8858	0.6182	0.8784
$\hat{g}_2(x_1z)$	AMSE	0.2019	0.8302	0.1388	0.7814	0.0973	0.7742
	ABIAS	0.1369	0.5877	0.0949	0.5871	0.0670	0.5987
	ASD	0.2428	0.9690	0.1991	0.9329	0.1849	0.9306
<i>Panel C</i>		$n = 100$		200		400	
	$(T = 5)$	FE	Pool	FE	Pool	FE	Pool
$\hat{m}_1(x_1)$	RAMSE	0.1801	0.8294	0.1245	0.8363	0.0833	0.8559
	ABIAS	-0.1182	-0.5759	-0.0828	-0.5908	-0.0559	-0.6113
	ASD	0.4918	0.5097	0.4605	0.4975	0.4289	0.5052
$\hat{m}_2(x_2)$	RAMSE	0.1347	0.4437	0.0933	0.4047	0.0634	0.3906
	ABIAS	0.0920	0.2818	0.0638	0.2555	0.0433	0.2499
	ASD	0.4050	3.8577	0.4044	3.8443	0.4032	3.8253
$\hat{f}(z)$	RAMSE	0.6291	12.0171	0.3938	9.2310	0.2778	6.0768
	ABIAS	-0.3676	-7.5230	-0.2346	-5.4599	-0.1661	-3.6358
	ASD	0.7075	12.6450	0.4873	9.9936	0.3969	6.7136
$\hat{g}_1(x_1x_2)$	RAMSE	0.1868	0.7438	0.1293	0.7082	0.0887	0.7055
	ABIAS	-0.1284	-0.5317	-0.0888	-0.5318	-0.0606	-0.5494
	ASD	0.6443	0.8803	0.6257	0.8456	0.6140	0.8371
$\hat{g}_2(x_1z)$	RAMSE	0.1759	0.7876	0.1217	0.7496	0.0852	0.7346
	ABIAS	0.1188	0.5646	0.0836	0.5633	0.0582	0.5733
	ASD	0.2263	0.9296	0.1932	0.9016	0.1802	0.8926



Table 3.4: Nonparametric Tests Results ( $T = 3$ )

		Test $\hat{T}_{n1}$ under $H_{01}$						Test $\hat{I}_{n2}$ under $H_{02}$					
		Empirical Size			Empirical Power			Empirical Size			Empirical Power		
	$a$	$n=100$	200	400	$n=100$	200	400	$n=100$	200	400	$n=100$	200	400
$c_0 = 0$	0.01	0.007	0.009	0.011	1	1	1	0.016	0.013	0.008	0.379	0.617	0.818
	0.05	0.039	0.044	0.048	1	1	1	0.058	0.053	0.048	0.429	0.774	0.995
	0.1	0.08	0.099	0.101	1	1	1	0.167	0.128	0.104	0.649	0.894	1
$c_0 = 1$	0.01	0.007	0.008	0.011	1	1	1	0.038	0.022	0.016	0.384	0.604	0.826
	0.05	0.041	0.048	0.051	1	1	1	0.067	0.057	0.054	0.424	0.798	0.912
	0.1	0.081	0.094	0.102	1	1	1	0.184	0.123	0.106	0.661	0.916	1
$c_0 = 2$	0.01	0.004	0.007	0.009	1	1	1	0.019	0.013	0.007	0.308	0.602	0.813
	0.05	0.045	0.048	0.050	1	1	1	0.078	0.064	0.055	0.464	0.782	0.911
	0.1	0.093	0.097	0.101	1	1	1	0.189	0.133	0.092	0.685	0.998	1

$\hat{f}(z)$ ). The Pool estimates are in general consistent under the random effect model, except for  $\hat{m}_1(x_1)$ ,  $\hat{g}_2(x_1z)$ , and  $\hat{g}_2(x_1z)$  even with a fairly large sample size (see *Panel A-C* of Table 3.2 when  $n = 400$ ). We also observe from Table 3.2 that the Pool estimator converges in a much slower rate than that of FE estimator. Under a SF-FE model in Table 3.3, however, the Pool estimator is largely biased and inconsistent. Thus, a comparison of the performance between FE and Pool reveals a considerable gain by incorporating the fixed effect in a SF model, which can be substantial to improve the reliability of both frontier and inefficiency estimates. Overall, our FE estimator demonstrates its appealing performance, which are robust to SF-RE and SF-FE when the regressors are fairly correlated to each other.

Last, we investigate the performance of our proposed test statistics  $\hat{T}_{n1}$  and  $\hat{I}_{n2}$  in terms of their empirical size and power. For  $\hat{T}_{n1}$  in (3.15), we test  $H_{01}: \Lambda(Q_{it}) = \Lambda(Q_{it}; \eta)$ , where  $\Lambda(Q_{it}; \eta) = \alpha + Q_{it}^\top \eta$  from

$$\text{DGP}_{01}: y_{it} = \alpha_i + \alpha_0 + Q_{it}^\top \eta + v_{it} - u_{it}$$

where  $\alpha_0 = 1$  is invariant over  $i$  and  $t$ ,  $Q_{it} = [x_{1,it}, x_{2,it}, z_{it}, \zeta_{it}^\top]^\top$ ,  $\zeta_{it} = [x_{1,it}x_{2,it}, x_{1,it}z_{it}]^\top$ , and  $\eta = [\eta_{x_1}, \eta_{x_2}, \eta_z, \eta_{x_1x_2}, \eta_{x_1z}]^\top = [1, 1.5, -1, 2, 0.5]^\top$ . We maintain the specification on the inefficiency function  $\mu(w_{it}; \gamma)$ , the distribution of variables  $\{x_{1,it}, x_{2,it}, z_{it}, w_{1,it}\}_{i=1, t=1}^{n, T}$ , and  $\alpha_i$  as in (3.18). We estimate  $\eta \equiv \eta(\gamma_0)$  as discussed in Section 2.2, and consistently estimate  $\alpha_0$  by  $\hat{\alpha}_0 = \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T (\hat{y}_{it}(\hat{\gamma}_0) - Q_{it}^\top \hat{\eta}(\hat{\gamma}_0))$ . We use DGP in (3.18) as the alternative of  $H_{01}$ . For  $\hat{I}_{n2}$  in (3.17), we test  $H_{02}: g_2(x_1z) = 0$  in (3.18), so the null model is given by:

$$\text{DGP}_{02}: y_{it} = \alpha_i + m_1(x_{1,it}) + m_2(x_{2,it}) + f(z_{it}) + g_1(x_{1,it}x_{2,it}) + v_{it} - u_{it},$$

We approximate the null distribution of both tests via the wild-bootstrap procedures outlined in Section 2.2. For  $\hat{T}_{n1}$ , we adopt a *rule of thumb* bandwidth  $h = [h_x, h_z]$  in (3.15) from  $h = c\hat{\sigma}(nT)^{-\frac{1}{7}}$ , where  $\hat{\sigma}$  is the empirical standard deviation of  $\{x_{1,it}, x_{2,it}, z_{it}\}_{i=1, t=1}^{n, T}$ , and  $c = 1$  is used in our simulation.

For  $\hat{I}_{n2}$ , we select  $J$  based on the same selection procedure as in (3.18). We obtain the empirical size and power as the average of rejection frequency from 1000 repetition with  $B = 399$  bootstrap. We further investigate the impact of fixed effect on the test performance by changing  $c_0 = (0, 1, 2)$ . In each test, we set the sample size  $n = (100, 200, 400)$  with  $T = 3$  for brevity.<sup>5</sup>

Table 3.4 shows the testing results of  $\hat{T}_{n1}$  on the left panel and  $\hat{I}_{n2}$  on the right panel. The empirical size and power are obtained for significant level  $\alpha = (0.01, 0.05, 0.10)$ , and reported beneath each size of  $n$ . We observe that the empirical size of  $\hat{T}_{n1}$  ( $\hat{I}_{n2}$ ) is undersized (oversized) when  $n = 100$ , but approaches steadily to its corresponding nominal level as  $n$  doubles. The empirical powers of  $\hat{T}_{n1}$  are all 1 across different  $n$ ,  $\alpha$ , and  $c_0$ . The empirical powers of  $\hat{I}_{n2}$  are moderately lower with small sample size  $nT = 300$ , but increases rapidly toward unity when  $n$  reaches 400. Overall, both tests exhibit reasonably well size and power across both Sf-RE and SF-FE model.

### 3.3 Empirical Application

We illustrate the empirical applicability of the SF-AMIFE through an application on a world production SF model. A number of studies have estimated world production model under SF framework, albeit with model setup. Kumbhakar and Wang (2005a) estimate the world production frontier with panel data for 82 countries during 1960-1987. Their study model the determinants of inefficiency as a function of time trend, and did not consider the role of environment variables. Pires and Garcia (2012) adopt the SF model specification as in Kumbhakar and Wang (2005a) with a long panel of 75 countries during 1950-2000. Employing a pooled estimator, Iyer et al. (2008) estimate the SF model in 20 OECD countries during 1982-2000, which incorporates human capital and R&D as environment variable in the frontier, and several trade variables, such as trade openness, in the inefficiency mean function. However, all studies above require a parametric specification *a priori* on both frontier and composite error distribution. Recently, Zhang et al. (2018) specify a SF-SC model to investigate the effect of economic freedom, an environment variable, on the frontier and inefficiency using 110 countries during 1990-2010. Similarly, Yao et al. (2018a) employ a SF-SC model with human capital as an environment variable using 134 countries during 1990-2011. However, both SF-SC models only include one and different environment variable at a time, and do not consider other potential determinants of inefficiency.<sup>6</sup> The above SF models also did not control for country-specific fixed effect.

Therefore, it is interesting to estimate the stochastic world production model by SF-AMIFE to incorporate important determinants of both frontier and inefficiency addressed in the literature. We source

<sup>5</sup>The results with  $T = 4$  and  $T = 5$  remain qualitatively similar. We thus omit them to save space.

<sup>6</sup>Under a SF-SC model, expanding the size of either environment variables or determinants of inefficiency will introduce the *curse of dimensionality*.

our aggregate country-level data from version 9.0 of Penn World Table (PWT), and merge it with *Economic Freedom of the World: 2015 Annual Report* (EFW) by Gwartney (2015) for the proxy of economic freedom. We use the PWT dataset because it provides the most comprehensive country-level panel data for our variables with reliable measures. The PWT also constructs country's gross domestic product (GDP) by its output-side real value in million of 2011 U.S. dollars, which can facilitate the comparison of economic productivity across countries and time. The EFW has been widely used in the literature, providing us the economic freedom index (based on a 0-10 scale) with higher score representing higher degree of economic freedom. Thus, a combination of the two datasets allow us to consider several interested variables in our application. To ensure a relatively large size of our sample in recent years, we extract a balanced panel of 116 countries during 2000-2013 to estimate the following SF-AMIFE:

$$\begin{aligned}
 Y_{it} = & \alpha_i + m_1(K_{it}) + m_2(L_{it}) + f_1(EF_{it}) + f_2(H_{it}) \\
 & + g_1(K_{it}L_{it}) + g_2(K_{it}EF_{it}) + g_3(L_{it}EF_{it}) \\
 & + g_4(K_{it}H_{it}) + g_5(L_{it}H_{it}) + g_6(EF_{it}H_{it}) + v_{it} - u_{it},
 \end{aligned} \tag{3.19}$$

where for  $n = 1, \dots, 116$  and  $t = 1, \dots, 14$ ,  $Y_{it}$  is measured by the output-side real GDP,  $K_{it}$  real capital stock,  $L_{it}$  the millions of persons engaged in employment,  $H_{it}$  the human capital index based on years of schooling and returns to education (Barro and Lee, 2013), and  $EF_{it}$  the chain-linked economic freedom index for its comparability across time. All  $Y$ ,  $K$ ,  $L$ , and  $H$  are taken natural log, with  $Y$  and  $K$  measured in millions of 2011 U.S. dollars. Thus, we allow two traditional production inputs ( $K, L$ ) to nonlinearly shift the frontier through  $(m_1(\cdot), m_2(\cdot))$ , possibly in a concave fashion. We also model the two inputs interaction effect through function  $g_1(\cdot)$ , the slope of which depends on their elasticity of substitution. Furthermore, we follow Zhang et al. (2018) and Yao et al. (2018a) to model the neutral effect of both  $EF$  and  $H$  through  $(f_1(\cdot), f_2(\cdot))$ , which are potentially nonlinear. We also follow them to incorporate the interaction effects between  $(K, L)$  and  $EF$  in functions  $(g_2(\cdot), g_3(\cdot))$ , and between  $(K, L)$  and  $H$  in functions  $(g_4(\cdot), g_5(\cdot))$ . We reflect the potential interaction of  $EF$  and  $H$  in  $g_6(\cdot)$ , as the neutral impact of human capital may be facilitated further in a country with higher economic freedom.

We also include a set of determinants of inefficiency addressed in the literature. First, a higher  $H$  can significantly improve country's productivity, possibly through enhancing technical efficiency (Miller and Upadhyay, 2000; Kumbhakar and Lovell, 2000). An increase in  $H$  also reduces the relative inefficiency among OECD countries (Kneller and Stevens, 2006). Second, countries with higher  $EF$  can improve its technical efficiency (Adkins et al., 2002; Klein and Luu, 2003; Zhang et al., 2018). Third, international trade can influence country's inefficiency through the effects of exports or imports. On one hand, engaging

Table 3.5: Data Descriptive Summary Statistics

Variables	Notation	Mean	S.D.	Min	$Q_{25}$	$Q_{75}$	Max
$\ln$ Real GDP	$Y$	11.694	1.844	7.282	10.169	12.97	16.6
$\ln$ Capital	$K$	12.737	1.965	7.416	11.249	14.165	17.934
$\ln$ Labor	$L$	1.657	1.591	-2.528	0.724	2.733	6.676
$\ln$ Human capital	$H$	0.887	0.287	0.112	0.699	1.12	1.315
Economic freedom index	$EF$	6.792	0.984	2.88	6.13	7.52	9.07
Export share	$EX$	0.307	0.289	0.013	0.114	0.431	2.086
Import share	$IM$	0.35	0.289	0.013	0.154	0.454	2.425
(n, T)	(116,14)						

in exports leads to international competition, which improves technical efficiency by reducing managerial inefficiency and resources mis-allocation (Chen and Tang, 1987; Egan and Mody, 1992; Clerides et al., 1998). On the other hand, import competition may impede the development of domestic firms in infancy (Shafaeddin, 2005). Countries may also experience higher efficiency over time through technological progress. We incorporate the above factors as the determinants of inefficiency in inefficiency mean function  $\mu(w_{it}; \gamma)$  from

$$\mu(w_{it}; \gamma) = \exp\{\gamma_0 + EF_{it}\gamma_{EF} + H_{it}\gamma_H + EX_{it}\gamma_{EX} + IM_{it}\gamma_{IM} + t\gamma_T\}, \quad (3.20)$$

where  $\gamma_0$  is a constant,  $EX_{it}$  ( $IM_{it}$ ) is exports (imports) share of GDP, measured by share of merchandise exports (imports) at current PPPs, and  $t$  is the linear time trend to capture the effect of technological progress. Finally, country-specific unobserved heterogeneity is likely to present in our study and correlated with our regressors. For instance, countries with coastline or trade-driven economic growth strategy may be more inclined to engage in trade. We use  $\alpha_i$  in (3.19) to capture the fixed effect. In our estimation, we follow the insights from simulation study to select  $J$ . Table 3.5 provides the data descriptive summary, including the first and third quantile ( $Q_{25}$ ,  $Q_{75}$ ) of each variables. See Feenstra et al. (2015) and Gwartney (2015) for a detailed data description.

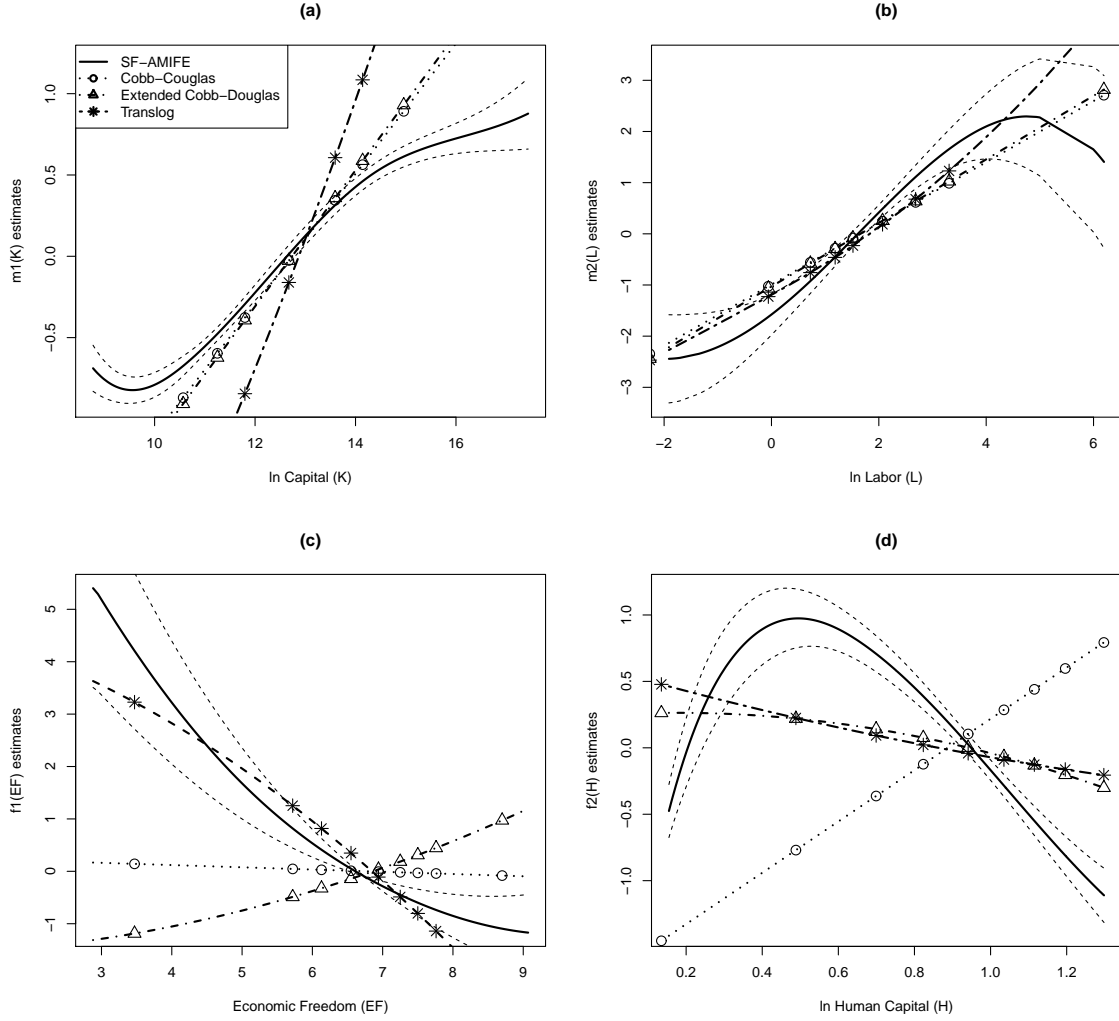
To be consistent with the literature on the SF model, we further compare our SF-AMIFE in (3.19) with three popular parametric SF models in the literature. The first model is a Cobb-Douglas SF model (SF-CD):

$$Y_{it} = \alpha_0 + \alpha_i + K_{it}\beta_K + L_{it}\beta_L + EF_{it}\beta_{EF} + H_{it}\beta_H + v_{it} - u_{it}, \quad (3.21)$$

where the neutral functions of both inputs and environment variables are linearly without interactions. The second model considered is a Extended Cobb-Douglas SF model (SF-ECD):

$$Y_{it} = \alpha_0 + \alpha_i + K_{it}\beta_K + L_{it}\beta_L + EF_{it}\beta_{EF} + EF_{it}^2\beta_{EF2} + H_{it}\beta_H + H_{it}^2\beta_{H2} + v_{it} - u_{it}, \quad (3.22)$$

Figure 3.2: SF-AMIFE: Neutral Functions Comparison



where we allow  $H$  and  $EF$  to neutrally shift the frontier quadratically, while maintain the linearity of inputs. The third parametric SF model is a Translog SF model (SF-TL):

$$\begin{aligned}
 Y_{it} = & \alpha_0 + \alpha_i + K_{it}\beta_K + K_{it}^2\beta_{K2} + L_{it}\beta_L + L_{it}^2\beta_{L2} + EF_{it}\beta_{EF} + EF_{it}^2\beta_{EF2} \\
 & + H_{it}\beta_H + H_{it}^2\beta_{H2} + K_{it}L_{it}\beta_{KL} + K_{it}EF_{it}\beta_{K,EF} \\
 & + L_{it}EF_{it}\beta_{L,EF} + K_{it}H_{it}\beta_{K,H} + L_{it}H_{it}\beta_{L,H} + EF_{it}H_{it}\beta_{EF,H} + v_{it} - u_{it},
 \end{aligned} \tag{3.23}$$

where both inputs and environment variables can impact the frontier through a quadratic function, with their pair-wise interaction captured by their product values. All three parametric SF models above have  $\mu(w_{it}; \gamma)$  defined in (3.20). Clearly, our SF-AMIFE nests the three parametric SF models as special cases.

We first compare the estimated neutral functions by plotting  $\hat{m}_1(\cdot)$ ,  $\hat{m}_2(\cdot)$ ,  $\hat{f}_1(\cdot)$ , and  $\hat{f}_2(\cdot)$  in Panel

Table 3.6: OLS FE Estimation Results

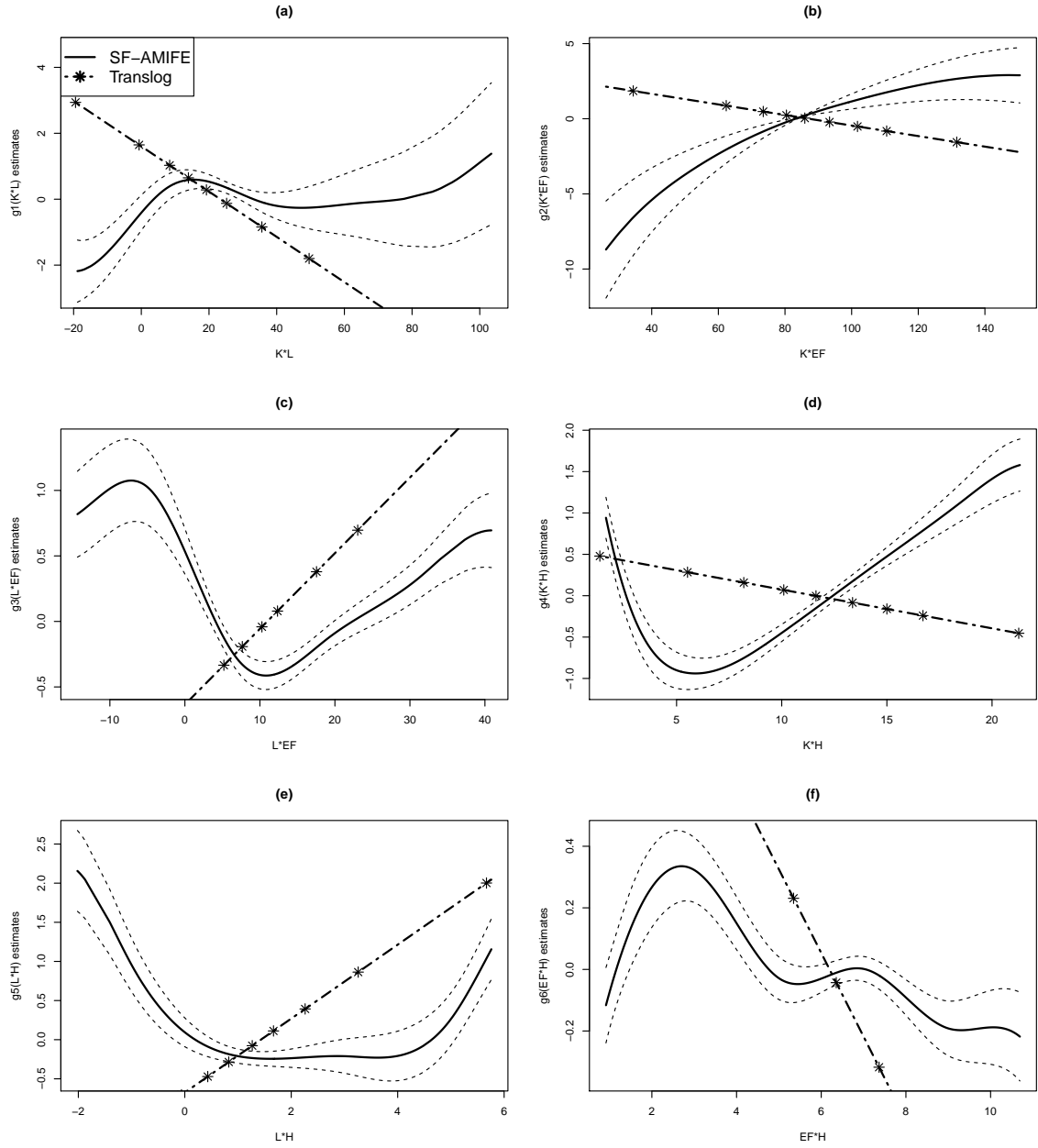
<i>Panel A</i>	variables	Cobb-Douglas	Extended Cobb-Douglas	Translog
	$\beta_0$	5.8439*** (0.0087)	5.2716*** (0.0074)	7.9377*** (0.0515)
	$K$	0.4011*** (0.0093)	0.4240*** (0.0099)	0.3896*** (0.0091)
	$K^2$	—	—	0.0283*** (0.0097)
	$L$	0.5974*** (0.0102)	0.5884*** (0.0108)	0.5547*** (0.1399)
	$L^2$	—	—	0.0551*** (0.0121)
	$EF$	-0.0428*** (0.0096)	-0.1893*** (0.0689)	-0.2456*** (0.1048)
	$EF^2$	—	0.0172*** 0.0054	-0.0688*** (0.0080)
	$H$	-1.9316*** (0.0427)	1.1744*** (0.1512)	-0.8121* (0.4575)
	$H^2$	—	-0.5816*** (0.0891)	0.1569 (0.1988)
	$K * L$	—	—	-0.0687*** (0.0208)
	$K * EF$	—	—	-0.0351*** (0.0106)
	$L * EF$	—	—	0.05776*** (0.0126)
	$K * H$	—	—	-0.0468 (0.0734)
	$L * H$	—	—	0.4723*** (0.0860)
	$EF * H$	—	—	-0.2717*** (0.0673)
<i>Panel B</i>	$\hat{\gamma}_{EF}$	-0.0859*** (0.0157)	-0.6518*** (0.0155)	-0.2354*** (0.0147)
	$\hat{\gamma}_H$	-1.525*** (0.0422)	2.8068*** (0.0435)	-1.0618*** (0.0614)
	$\hat{\gamma}_{EX}$	-0.8922*** (0.0357)	2.0391*** (0.0357)	-0.2201*** (0.0372)
	$\hat{\gamma}_{IM}$	1.1819*** (0.0211)	-7.6633*** (0.0214)	0.3331*** (0.0134)
	$\hat{\gamma}_T$	0.0037 (0.0147)	-0.233*** (0.0198)	-0.0033 (0.0345)
	NLS Score	21.5325	21.1579	19.793
	FE	YES	YES	YES
	Country (n)	116	116	116
	Time (T)	14	14	14
	Total Obs:	1624	1624	1624

(a)-(d) of Figure 3.2, respectively, by SF-AMIFE (solid line) with 95% wild-bootstrap CI (thin dash line), SF-CD (dot line with circle), SF-ECD (short dash-dot line with triangle), and SF-TL (long dash-dot line with star). Due to the identification condition, our neutral functions need to be centered. For comparison purpose, we also center the estimated parametric functions, with their corresponding estimates reported in *Panel A* of Table 3.6. We also estimate the derivative of each neutral functions in SF-AMIFE. For brevity, we will summarize the estimated derivatives of SF-AMIFE throughout the discussion below.

In Panel (a)-(b) of Figure 3.2, our SF-AMIFE reveals that both  $K$  and  $L$  shift upward the frontier nonlinearly, with its diminishing return clearly observed at the upper boundary. This is consistent with the economic theory regarding the concavity of the production inputs. In contrast, SF-CD and SF-ECD slightly overestimate (underestimate) the slope of  $m_1(K)$  ( $m_2(L)$ ), and SF-TL predicts a convex, rather than concave, shape of both functions ( $(\hat{\beta}_K, \hat{\beta}_{K2}) = (0.3896, 0.0283)$ ,  $(\hat{\beta}_L, \hat{\beta}_{L2}) = (0.5547, 0.0551)$ ). Our SF-AMIFE gives the mean of derivative estimates  $m'_1(K)$  ( $m'_2(L)$ ) by 0.3313 (0.6368), relatively lower (higher) than the parametric estimates. Panel (c) shows that SF-AMIFE disclose a declining trend in the neutral function of  $EF$ , with its marginal effects also decreased as  $EF$  rises. The derivative estimates  $f'_1(EF)$  reduces from -0.926 to -0.562 as  $EF$  increases from its first to third quantile, with an average of -0.764. Thus, a higher  $EF$  exhibits a lower facilitating impact of  $EF$  on the frontier. This is consistent with Yao et al. (2018b), showing that an increase in  $EF$  relaxes governmental control of population, inducing many immigrants to migrate to other countries with higher  $EF$ . Thus, the original country may experience less increase in output. We observe that SF-CD and SF-ECD of  $\hat{f}_1(EF)$  largely deviate from the semiparametric estimates, and SF-TL gives  $\hat{f}_1(EF)$  with a different curvature. Panel (d) demonstrates a significant inverted U shape in the neutral function of  $\hat{f}_2(H)$  by SF-AMIFE, with the frontier improved with  $H < 0.5$  but declined afterwards with higher  $H$ . The declining part of  $\hat{f}_2(H)$  also in line with Yao et al. (2018a), documenting a diminishing return of  $H$  on country's output through its neutral function. However, all three parametric estimates largely fall outside of the 95% CI constructed for SF-AMIFE, thus failing to capture the nonlinearity of human capital neutral function.

We further compare the estimated interaction functions in Figure 3.3. In each panel (a)-(f), we plot the estimated  $\hat{g}_k(\cdot)$ ,  $k = 1, \dots, 6$ , by SF-AMIFE against the SF-TL estimates. One clear observation is that the parametric functions do not capture the majority shape of each interaction function, with some of them having opposite slope. Based on the SF-AMIFE estimates, panel (a) shows that  $K$  and  $L$  jointly increases the frontier through  $\hat{g}_1(KL)$  when  $K * L$  are kept less than 18. Given the positive output elasticity of  $K$  and  $L$  in panel (a)-(b) of Figure 3.2, we observe a complementary effect among inputs when  $KL < 18$ , but a more substitutable effect when  $18 < KL < 45$ . This can be potentially due to the capital-biased technology, enlarging the capital-labor elasticity of substitution (Blanchard et al.,

Figure 3.3: SF-AMIFE: Interaction Functions Comparison





1997). Similar to the finding in Zhang et al. (2018) and Yao et al. (2018a), an increasing  $EF$  or  $H$  rises the frontier by improving the output elasticity of  $K$  through (panel (b), (d)), but decreases the frontier through lowering the output elasticity of labor (panel (c), (e)) with  $L * EF < 10$  and  $L * H < 4.2$ . Panel (f) shows that  $EF$  and  $H$  jointly increases  $Y$  through  $\hat{g}_6(EF * H)$  with a fairly low level of  $EF * H < 3$ , followed by steadily declining trend as  $EF * H$  further rises.

Finally, we evaluate the partial effect of  $K$ ,  $L$ ,  $EF$ , and  $H$  on the frontier. We follow the formula (3.10) to compute  $dY/dK$  and  $dY/dL$ , and (3.11) to compute  $dY/dEF$  and  $dY/dH$ . Since partial effect of each variables depends all  $(K, L, EF, H)$  and derivate estimates, we summarize each measures below based on their mean value. We observe that output elasticity of  $K$  ( $dY/dK$ ) in SF-SMIFE has a mean of 0.393, fairly close to its parametric counterparts in SF-CD (0.401), SF-ECD (0.424), and SF-TL (a mean of 0.417). However, the output elasticity of labor ( $dY/dL$ ) has a mean of 0.768 in SF-AMIFE, much higher than its mean of 0.674 in SF-TL and both SF-CD (0.597) and SF-ECD (0.589). *EconomicfreedomEF* in SF-AMIFE improves the frontier by a marginal impact of 0.204 on average. However, all parametric models predicts a negative partial effect of  $EF$ , with the SF-TL giving the largest estimates of -1.773, followed by SF-ECD of -0.044 and SF-CD of -0.043.  $H$  in SF-SMIFE promotes the frontier marginally by 0.041 on average, but declines the frontier by a magnitude of more than 1.5 across parametric models. Hence, our SF-AMIFE provides a more reliable and intuitive measures on the elasticity of inputs and environment variables.

In sum, our SF-AMIFE reveals significant nonlinear effects of inputs, environment variables, and their joint effects, all of which are not captured adequately by their parametric SF models. Thus, we conjecture that the parametric models may suffer from model misspecification. We test the conjecture by implementing our test  $T_{n1}$  with 399 bootstrap repetition, with the null model structure being SF-CD, SF-ECD, or SF-TL. The test results are given in the upper panel of Table 3.4 with the four-digit zero empirical p-values under each null. Thus, we reject all parametric models at 1%, indicating that the conventional specification on the frontier can be still restrictive to reveal the complexity of production frontier model. We further provide evidence on the presence of each interaction function  $g_{1-6}(\cdot)$  by testing  $H_{02} : g_k(\cdot) = 0$  for  $k = 1, \dots, 6$ . The results are reported in lower panel of Table 3.4. We find a weak evidence on the interaction of  $K$  and  $L$ , which is significant at 10% level. The joint effect of  $g_2(K * EF)$ ,  $g_2(L * EF)$ ,  $g_2(K * H)$ ,  $g_2(K * L)$  are all highly significant, again in line with Zhang et al. (2018) and Yao et al. (2018a). The interaction effect between  $EF$  and  $H$  is also significant at 1% level, highlighting the importance of modeling them in empirical studies.

Regarding the inefficiency mean function, the SF-AMIFE has the estimates

$$\hat{\gamma} = (\hat{\gamma}_0, \hat{\gamma}_{EF}, \hat{\gamma}_H, \hat{\gamma}_{EX}, \hat{\gamma}_{IM}, \hat{\gamma}_T) = (0.1123, -0.1406, -1.4721, -0.5383, 0.7832, -0.0191), \text{ with standard}$$

Table 3.7: Nonparametric Tests Results on Frontier Functional Form and Interactions

Model Specification Test $H_{01}$ : OLS FE		
OLS Model	$\hat{T}_{n1}$	p-value
CD	5.6649	0.0000
ECD	4.2147	0.0000
TL	5.3948	0.0000
Presence of Interaction Test $H_{02}$ : $g_k(\cdot) = 0$		
Interaction	$\hat{I}_{n2}$	p-value
$g_1(K * L)$	0.4810	0.0920
$g_2(K * EF)$	2.9838	0.0175
$g_3(L * EF)$	1.1950	0.0000
$g_4(K * H)$	1.4821	0.0000
$g_5(L * H)$	0.2440	0.0125
$g_6(EF * H)$	1.6269	0.0000

error given by (0.0321, 0.0802, 0.0446, 0.0615, 0.1848, 0.0032). Here, a positive sign indicates a positive effect on the *inefficiency* mean, thus a negative effect on the productive *efficiency*. Consistent with the finding in the literature, we find that countries endowed with higher human capital and level of economic freedom tend to be more efficient. Engaging in exports also contributes to lower inefficiency in a country due to the international trade competition, although the domestic competition induced by imports enlarges the inefficiency level. Finally, countries are more efficient over time due to the technological progress. The parametric estimates of  $\hat{\gamma}$  in *Panel B* of Table 3.6 have the sign qualitatively similar to that of the SF-AMIFE, except some counter-intuitive estimates of  $\gamma_H = 2.807$ ,  $\gamma_{EX} = 2.039$ , and  $\gamma_{IM} = -7.6633$  by SF-ECD.

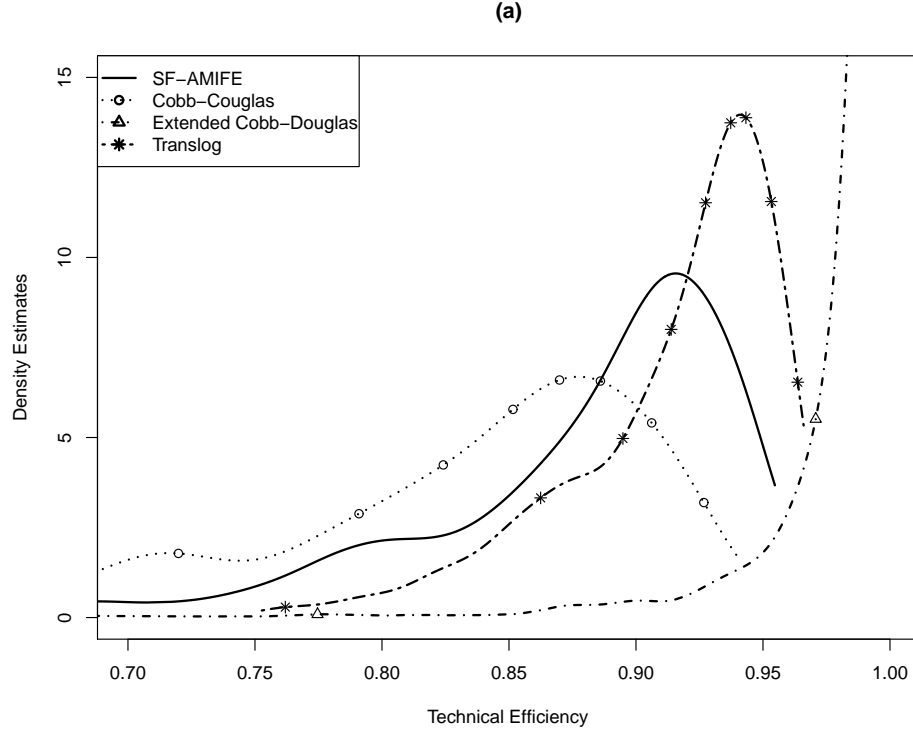
Our estimates  $\hat{\gamma}$  also allows us to construct country-specific technical efficiency (TE) level, which can be of interest to regulators. One popular approach is to compute  $TE = \exp\{-\hat{u}_{it}\}$ , where  $\hat{u}_{it} = \hat{E}(u_{it}|\epsilon_{it})$  is computed based on the convolution density of  $v_{it} - u_{it}$  by Jondrow et al. (1982) (JLMS). However, the JLMS requires the distribution of the composite error to be specified. We propose to construct  $TE = \exp\{-\hat{\mu}(w_{it}; \hat{\gamma})\}$ , where  $\hat{\gamma}$  is obtained from SF-AMIFE or the parametric models. To provide a vivid picture of our TE estimates, we plot its density across SF-AMIFE, SF-CD, SF-ECD, and SF-TL in panel (a) of Figure 3.4. The TE from SF-AMIFE has a mean of 0.8789, clearly revealing a relatively large fraction of inefficient countries. Comparing with SF-AMIFE, the TE is underestimated by SF-CD with a mean of 0.8289, but largely overestimated in SF-ECD (0.956) and SF-TL (0.914). Thus, a misspecified frontier model can lead to a misleading conclusion regarding the level of technical efficiency across countries. We further calculate the yearly average of TE (ATE) and of the change of TE (ATEC) during 2001-2013, and report the top and bottom 20 countries in our sample based on the ranking of

Table 3.8: Yearly Averaged TE and TEC Ranking 2001-2013

	Country Name	ATE	ATEC
Top 20 Countries	Norway*	0.9518	0.0005
	United States*	0.9509	-0.0002
	Switzerland*	0.9499	0.0001
	Canada*	0.9498	0.0000
	Australia*	0.9486	-0.0002
	United Kingdom*	0.9476	0.0000
	Germany*	0.9468	0.0000
	Japan*	0.9455	-0.0001
	New Zealand*	0.9421	-0.0002
	Denmark*	0.9412	0.0000
	Ireland*	0.9406	0.0006
	Czech Republic*	0.9386	0.0001
	Finland*	0.9382	-0.0003
	Israel*	0.9377	0.0010
	Republic of Korea	0.9374	0.0008
	Sweden*	0.9361	-0.0001
	Netherlands*	0.9339	0.0000
	Estonia*	0.9327	0.0000
	Slovakia*	0.9317	0.0005
	Austria*	0.9315	-0.0003
Bottom 20 Countries	Niger	0.6712	0.0045
	Burundi	0.6885	0.0073
	Mali	0.6988	0.0055
	Senegal	0.7335	0.0037
	Myanmar	0.7601	0.009
	Central African Republic	0.7634	0.0035
	Sierra Leone	0.7664	0.0089
	Benin	0.7673	0.0051
	Nepal	0.7776	0.0057
	D.R. of the Congo	0.7803	0.0054
	Cte d'Ivoire	0.7829	0.0018
	Rwanda	0.7907	0.0084
	Malawi	0.7908	0.0037
	Madagascar	0.7917	0.0036
	U.R. of Tanzania: Mainland	0.7969	0.0031
	Haiti	0.8006	0.0007
	Morocco	0.8056	0.0033
	Togo	0.8101	0.0039
	Pakistan	0.8182	0.0041
	Nigeria	0.8236	0.0042

*Note:* Countries with star are OECD countries by 2013.

Figure 3.4: SF-AMIFE: Interaction Functions Comparison



TE in Table 3.8. All countries but South Korea are OECD countries (with star) in the Top 20 list, with Norway, United States, and Switzerland as the top three efficient countries in the world. In contrast, most countries in the Bottom 20 list are Africa countries, with Niger, Burunidi, and Mali the bottom three inefficient countries. The average of ATE over the Top 20 countries is 0.942, about 22% higher than the mean of ATE over the bottom 20 countries (0.771). Similar to the finding in Iyer et al. (2008), ATEC is moderately lower on the Top 20 than the Bottom 20. This is expected, since high efficient countries can produce very near to the frontier, thus not likely to experience a further notable improvement (i.e., a slow change of TE). However, inefficient countries may catch up with their frontier more quickly by exploring available opportunities that effectively reduce production inefficiency, such as accumulating human capital. Overall, our SF-AMIFE provides a fairly reasonable measure on the TE and as well as its ranking.

### 3.4 Conclusion

In this paper, we propose a semiparametric additive stochastic frontier model with several features that adds to the literature. We model the frontier as an additive unknown functions of transitional inputs, environment variables, as well as their possible interactions, all of which are nonparametrically estimated

to capture the potential nonlinearities of interests. We allow the mean function of inefficiency to be influenced by a set of its determinants, which may or may not appear on the frontier model. To avoid mis-measuring the inefficiency level, we adopt a panel data structure by separating out the time-invariant fixed effect from inefficiency. We estimate the model consistently through B-spline estimator combined with NLS, which does not require restrictive assumption on the composite error distribution. Our estimator is computationally attractive, and does not suffer from the *curse of dimensionality* and incidental parameter problem. Hence, the environment variables and inefficiency determinants can be of high dimension in our model with a large panel data.

We demonstrate through simulation studies the finite sample performance of our proposed estimator, which works reasonably well under either random or fixed effect SF model with fairly correlated regressors. We also construct two nonparametric tests for the frontier functional form specification and the presence of interaction functions, both exhibiting reliable size and power. To illustrate the practical use of our model with the proposed tests, we employ 116 countries during 2001–2013 to perform an world production frontier model estimation. Follow the recent literature, we include economic freedom and human capital as environment variables that influence the frontier with traditional production inputs, and also as inefficiency determinants together with trade variables and technological progress. We find that both inputs significantly rise the frontier with diminishing return, and both environment variables shift the frontier in a nonlinear fashion. The environment variables also collaborate with the inputs on shifting the frontier, consistent with the finding in recent studies. We show that the above nonlinear effects are not captured adequately by conventional parametric SF models, whose specification are all rejected 1% significant level. We observe that a country's technical efficiency is improved by human capital, economic freedom, exports, and technological progress, but lowered by import. Based on the ranking of TE over years, we see that OECD countries are much more efficient than Africa and Southern Asia countries.

# Bibliography

- Adkins, L. C., Moomaw, R. L., Savvides, A., 2002. Institutions, freedom, and technical efficiency. *Southern Economic Journal*, 92–108.
- Ai, C., You, J., Zhou, Y., 2014. Estimation of fixed effects panel data partially linear additive regression models. *The Econometrics Journal* 17 (1), 83–106.
- Aigner, D., Lovell, C. A. K., Schmidt, P., 1977. Formulation and estimation of stochastic frontiers production function models. *Journal of Econometrics* 6, 21–37.
- Barro, R. J., Lee, J. W., 2013. A new data set of educational attainment in the world, 1950–2010. *Journal of Development Economics* 104, 184–198.
- Battese, G. E., Coelli, T. J., 1992. Frontier production functions, technical efficiency and panel data: with application to paddy farmers in India. *Journal of Productivity Analysis* 3, 153–169.
- Blanchard, O. J., Nordhaus, W. D., Phelps, E. S., 1997. The medium run. *Brookings Papers on Economic Activity* 1997 (2), 89–158.
- Chen, T., Tang, D., 1987. Comparing technical efficiency between import-substitution-oriented and export-oriented foreign firms in a developing economy. *Journal of Development Economics* 26 (2), 277–289.
- Chen, Y., Schmidt, P., Wang, H., 2014. Consistent estimation of the fixed effects stochastic frontier model. *Journal of Econometrics* 181 (2), 65–76.
- Clerides, S. K., Lach, S., Tybout, J. R., 1998. Is learning by exporting important? micro-dynamic evidence from Colombia, Mexico, and Morocco. *The Quarterly Journal of Economics* 113 (3), 903–947.
- Colombi, R., Kumbhakar, S. C., Martini, G., Vittadini, G., 2014. Closed-skew normality in stochastic frontiers with individual effects and long/short-run efficiency. *Journal of Productivity Analysis* 42 (2), 123–136.
- Cornwell, C., Schmidt, P., Sickles, R. C., 1990. Production frontiers with cross-sectional and time-series variation in efficiency levels. *Journal of Econometrics* 46 (2), 185–200.
- Egan, M. L., Mody, A., 1992. Buyer-seller links in export development. *World Development* 20 (3), 321–334.
- Fan, Y., Li, Q., Weersink, A., 1996. Semiparametric estimation of stochastic production frontier models. *Journal of Business and Economic Statistics* 14, 460–468.
- Feenstra, R. C., Inklaar, R., Timmer, M. P., 2015. The Next Generation of the Penn World Table. *American Economic Review* 105, 3150–3182.
- Ferrara, G., Vidoli, F., 2017. Semiparametric stochastic frontier models: A generalized additive model approach. *European Journal of Operational Research* 258 (2), 761–777.

- Greene, W., 2005a. Fixed and random effects in stochastic frontier models. *Journal of Productivity Analysis* 23 (1), 7–32.
- Greene, W., 2005b. Reconsidering heterogeneity in panel data estimators of the stochastic frontier model. *Journal of Econometrics* 126 (2), 269–303.
- Greene, W. H., 1980. Maximum likelihood estimation of econometric frontier functions. *Journal of Econometrics* 13 (1), 27–56.
- Greene, W. H., 1993. The econometric approach to efficiency analysis. In: H. Fried, C. A. K. Lovell, and S. S. Schmidt, (Eds.), *The Measurement of Productive Efficiency*. Oxford University Press, Oxford.
- Gwartney, J., Lawson, R., Hall, J., 2015. *Economic freedom of the world: 2015 annual report*. Vancouver, Fraser Institute.
- Horrace, W. C., Parmeter, C. F., 2011. Semiparametric deconvolution with unknown error variance. *Journal of Productivity Analysis* 35 (2), 129–141.
- Iyer, K. G., Rambaldi, A. N., Tang, K. K., 2008. Efficiency externalities of trade and alternative forms of foreign investment in OECD countries. *Journal of Applied Econometrics* 23 (6), 749–766.
- Jondrow, J., Lovell, C. A. K., Materov, I. S., Schmidt, P., 1982. On the estimation of technical inefficiency in the stochastic frontier production function model. *Journal of Econometrics* 19, 233–238.
- Klein, P. G., Luu, H., 2003. Politics and productivity. *Economic Inquiry* 41 (3), 433–447.
- Kneip, A., Simar, L., Van Keilegom, I., 2015. Frontier estimation in the presence of measurement error with unknown variance. *Journal of Econometrics* 184 (2), 379–393.
- Kneller, R., Stevens, P. A., 2006. Frontier technology and absorptive capacity: Evidence from oecd manufacturing industries. *Oxford Bulletin of Economics and Statistics* 68 (1), 1–21.
- Kumbhakar, S. C., 1990. Production frontiers, panel data, and time-varying technical inefficiency. *Journal of Econometrics* 46 (1-2), 201–211.
- Kumbhakar, S. C., Lovell, C. A. K., 2000. *Stochastic frontier analysis*. Cambridge university Press, Cambridge, UK.
- Kumbhakar, S. C., Park, B. U., Simar, L., Tsionas, E., 2007. Nonparametric stochastic frontiers: a local maximum likelihood approach. *Journal of Econometrics* 137, 1–27.
- Kumbhakar, S. C., Parmeter, C. F., Zelenyuk, V., 2018. *Stochastic frontier analysis: Foundations and Advances*. Tech. rep.
- Kumbhakar, S. C., Wang, H.-J., 2005a. Estimation of growth convergence using a stochastic production frontier approach. *Economics Letters* 88 (3), 300–305.
- Kumbhakar, S. C., Wang, H.-J., Horncastle, A. P., 2015. *A Practitioner’s Guide to Stochastic Frontier Analysis Using STATA*. Cambridge University Press.
- Lee, Y., Schmidt, P., 1993. A production frontier model with flexible temporal variation in technical efficiency. In: in L. H. Fried & S. Schmidt, eds, *‘The Measurement of Productive Efficiency’*. Oxford University Press, Oxford, United Kingdom.
- Li, Q., 1996. Estimating a stochastic production frontier when the adjusted error is symmetric. *Economics Letters* 52 (3), 221–228.
- Li, Q., 2000. Efficient estimation of additive partially linear models. *International Economic Review* 41 (4), 1073–1092.
- Lin, Z., Li, Q., Sun, Y., 2014. A consistent nonparametric test of parametric regression functional form in fixed effects panel data models. *Journal of Econometrics* 178, 167–179.

- Linton, O. B., Nielsen, J. P., 1995. A kernel method of estimating structured nonparametric regression based on marginal integration. *Biometrika* 82, 93–100.
- Martins-Filho, C., Yao, F., 2015. Nonparametric stochastic frontier estimation via profile likelihood. *Econometric Reviews* 34 (4), 413–451.
- Meeusen, W., van Den Broeck, J., 1977. Efficiency estimation from Cobb-Douglas production functions with composed error. *International Economic Review* 18, 435–444.
- Miller, S. M., Upadhyay, M. P., 2000. The effects of openness, trade orientation, and human capital on total factor productivity. *Journal of Development Economics* 63 (2), 399–423.
- Park, B. U., Simar, L., Zelenyuk, V., 2015. Categorical data in local maximum likelihood: theory and applications to productivity analysis. *Journal of Productivity Analysis* 43, 199–214.
- Parmeter, C. F., Kumbhakar, S. C., 2014. Efficiency Analysis: A Primer on Recent Advances. *Foundations and Trends® in Econometrics* 7 (3-4), 191–385.
- Parmeter, C. F., Wang, H.-J., Kumbhakar, S. C., 2016. Nonparametric estimation of the determinants of inefficiency. *Journal of Productivity Analysis*.
- Pires, J. O., Garcia, F., 2012. Productivity of nations: a stochastic frontier approach to TFP decomposition. *Economics Research International* 2012.
- Schmidt, P., Sickles, R. C., 1984. Production frontiers and panel data. *Journal of Business & Economic Statistics* 2 (2), 367–374.
- Shafaeddin, S. M., 2005. Trade policy at the crossroads: The recent experience of developing countries. Palgrave Macmillan.
- Simar, L., Wilson, P., 2008. Statistical inference in nonparametric frontier models: recent developments and perspectives. In: Fried, H., Lovell, C. A. K., Schmidt, S. S. (Eds.), *The Measurement of Productive Efficiency*. Oxford University Press, Oxford, UK, Ch. Chapter 4.
- Sperlich, S., Tjøstheim, D., Yang, L., 2002. Nonparametric estimation and testing of interaction in additive models. *Econometric Theory* 18 (2), 197–251.
- Stevenson, R. E., 1980. Likelihood functions for generalized stochastic frontier estimation. *Journal of Econometrics* 13 (1), 57–66.
- Sun, K., Kumbhakar, S. C., 2013. Semiparametric smooth-coefficient stochastic frontier model. *Economics Letters* 120 (2), 305–309.
- Vidoli, F., Ferrara, G., 2015. Analyzing Italian citrus sector by semi-nonparametric frontier efficiency models. *Empirical economics* 49 (2), 641–658.
- Wang, H.-J., Ho, C.-W., 2010. Estimating fixed-effect panel stochastic frontier models by model transformation. *Journal of Econometrics* 157 (2), 286–296.
- Wang, L., Xue, L., 2015. Constrained polynomial spline estimation of monotone additive models. *Journal of Statistical Planning and Inference* 167, 27–40.
- Yao, F., Wang, T., Tian, J., Kumbhakar, S. C., 2018a. Estimation of a smooth coefficient zero-inefficiency panel stochastic frontier model: A semiparametric approach. *Economics Letters* 166, 25–30.
- Yao, F., Zhang, F., Kumbhakar, S. C., 2018b. Semiparametric smooth coefficient stochastic frontier model with panel data. *Journal of Business & Economic Statistics*, 1–17.
- Zhang, F., Hall, J., Yao, F., 2018. Does economic freedom affect the production frontier? a semiparametric approach with panel data. *Economic Inquiry* 56 (2), 1380–1395.
- Zhang, R., Sun, K., Delgado, M. S., Kumbhakar, S. C., 2012. Productivity in china's high technology industry: Regional heterogeneity and r&d. *Technological Forecasting and Social Change* 79 (1), 127–141.